

# DAS MITTELALTER

Perspektiven mediävistischer Forschung  
Zeitschrift des Mediävistikverbandes

2025 · 30/1

**Schnittstelle Mediävistik**  
**Kollaborationen der Mittelalterforschung im digitalen Zeitalter**

Herausgegeben von  
Luise Borek  
Karoline Döring  
Nora Ketschik  
Katharina Zeppezauer-Wachauer



HEIDELBERG  
UNIVERSITY PUBLISHING



# **DAS MITTELALTER**

Perspektiven mediävistischer Forschung  
Zeitschrift des Mediävistikverbandes

2025 · 30/1

**Schnittstelle Mediävistik  
Kollaborationen der Mittelalterforschung  
im digitalen Zeitalter**

Herausgegeben von  
Luise Borek, Karoline Döring, Nora Ketschik  
und Katharina Zeppezauer-Wachauer

### **Herausgeberin**

Prof. Dr. Isabelle Mandrella  
Katholisch-Theologische Fakultät, Philosophie und philosophische  
Grundfragen der Theologie, LMU München, Geschwister-Scholl-Platz 1,  
80539 München, isabelle.mandrella@lmu.de

### **Redaktion**

Dr. Magdalena Butz, Dr. Jan Glück  
redaktion-das-mittelalter@posteo.de

Logo des Mediävistikverbandes von Walter Wolf, Riedstadt, nach Hs Florenz,  
Biblioteca Medicea Laurenziana, Conv. sopp. 319, f. 90v

### **Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der  
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind  
im Internet über <https://dnb.dnb.de> abrufbar.



Diese Zeitschrift ist unter der Creative-Commons-Lizenz  
CC BY-SA 4.0 veröffentlicht. Die Umschlaggestaltung  
unterliegt der Creative-Commons-Lizenz CC BY-ND 4.0.

Publiziert bei Heidelberg University Publishing (heiUP), 2025

Universität Heidelberg / Universitätsbibliothek  
Heidelberg University Publishing (heiUP)  
Grabengasse 1, 69117 Heidelberg  
<https://heiup.uni-heidelberg.de>  
E-Mail: [ub@ub.uni-heidelberg.de](mailto:ub@ub.uni-heidelberg.de)

Die Online-Version dieser Publikation ist auf den Verlagswebseiten von  
Heidelberg University Publishing <https://heiup.uni-heidelberg.de> dauerhaft  
frei verfügbar (Open Access).  
doi: <https://doi.org/10.17885/heiup.mial.2025.1>

Text © 2025. Das Copyright der Texte liegt bei den jeweiligen Verfasser:innen.

ISSN 0949-0345  
eISSN 2196-6869

ISBN 978-3-96822-337-7 (Softcover)  
ISBN 978-3-96822-336-0 (PDF)

Die Zeitschrift ‚Das Mittelalter‘ ist ein Forum für die interdisziplinäre Mediävistik. Die einzelnen Hefte behandeln jeweils ein aktuelles Thema der Mittelalter-Forschung unter fächerübergreifender Perspektive. Daneben werden in einem berichtenden Teil die wesentlichen Ergebnisse, Diskussionen und Neuerscheinungen der verschiedenen Disziplinen in ihrer Bedeutung für die gesamte Mediävistik vorgestellt.

### **Peer reviewed**

### **Der Mediävistikverband e. V.**

#### **Präsidium**

Prof. Dr. Regina Toepfer (Präsidentin), Würzburg  
Prof. Dr. Michael Grünbart (Vizepräsident), Münster  
Prof. Dr. Albrecht Fuess (Schatzmeister), Marburg  
Prof. Dr. Isabelle Mandrella (Schriftführerin), München  
Prof. Dr. Wolfram Drews (Vertreter des Tagungsortes), Münster

#### **Beirat**

#### **Funktionsstellen**

Prof. Dr. Andrea Sieber (Mediävistik und Schule), Passau  
Prof. Dr. Ingrid Baumgärtner (Beihefte), Kassel  
Dr. Karoline Döring (Öffentlichkeitsarbeit), Salzburg  
Dr. Miriam Hathaway (Early Career), Mainz

#### **Fachvertretung**

Prof. Dr. Eva von Contzen (Anglistik), Freiburg  
Prof. Dr. Stephan Dusil (Rechtsgeschichte), Tübingen  
Prof. Dr. Juliane von Fircks (Kunstgeschichte), Jena  
Prof. Dr. Susanne A. Friede (Romanistik), Bochum  
Prof. Dr. Albrecht Fuess (Islamwissenschaft), Marburg  
Prof. Dr. Michael Grünbart (Byzantinistik), Münster  
Prof. Dr. Irene Holzer (Musikwissenschaft), München  
PD Dr. Christian Kaiser (Medizingeschichte), Bonn  
Prof. Dr. Andreas Lehnertz (Jüdische Studien), Trier  
Dr. Christine Magin (Grundwissenschaften), Greifswald  
Prof. Dr. Isabelle Mandrella (Philosophie), München  
Prof. Dr. Christoph Mauntel (Geschichte), Osnabrück  
Prof. Dr. Natascha Mehler (Mittelalterarchäologie), Tübingen  
Prof. Dr. Jonathan Reinert (Theologie), Reutlingen  
Prof. Dr. Bernd Roling (Mittelalter), Berlin  
Prof. Dr. Roland Scheel (Skandinavistik), Münster  
Prof. Dr. Gabriel Viehhauser (Germanistik und Digital Humanities), Wien  
Jun.-Prof. Dr. Matthias Weber (Geschichte), Bochum  
Prof. Dr. Julia Zimmermann (Germanistik), Graz

Anfragen und Mitteilungen bezüglich der Mitgliedschaft (Beitrittserklärungen, Änderungen von Adresse und Bankverbindung) richten Sie bitte an den Schatzmeister des Mediävistikverbandes:  
Prof. Dr. Albrecht Fuess, Fachgebiet Islamwissenschaft, Deutschhausstraße 12, 35032 Marburg; albrecht.fuess@uni-marburg.de.

# Inhaltsverzeichnis

- Luise Borek, Hannah Busch, Karoline Döring,  
Nora Ketschik und Katharina Zeppezauer-Wachauer
- 1 **Einleitung**  
Schnittstelle Mediävistik.  
Kollaborationen der Mittelalterforschung im digitalen Zeitalter
- Sabine Tittel
- 18 **Historisierte Ontologien für *Linguistic Linked Open Data*-Ressourcen des Mittelalters**
- Bernhard Bauer, Roman Bleier and Jakob Sonnberger
- 38 **Linked Open Data and Manuscript Descriptions**
- Michael Schonhardt
- 54 **Die digitale Edition als Schnittstelle**  
Editorische Workflows im KI-Zeitalter
- Dana Eichhorst, Laura Frank, Danah Tonne und Katharina H. Wendl
- 70 **Interdisziplinär erforscht**  
Das Phänomen der *Tagin* und *Otíjjot Meshunnot* als  
paläographische Besonderheit in mittelalterlichen Torarollen
- Isabella Nicka, Miriam Landkammer, Andreas Uhl,  
Michael Linortner and Johannes Schuiki
- 90 **Annotating Images**  
A Prerequisite for Interdisciplinary Research in Medieval Studies
- Elisabeth Gruber und Peter Färberböck
- 113 **ITEM**  
Eine Schnittstelle realienkundlicher Daten aus historischen  
Textquellen
- Matthias Weber
- 130 **Bischöfe im Graph**  
Zur Transformation großer Datenbestände in eine graphbasierte  
Umgebung
- Benjamin Hitz und Tobias Hodel
- 148 **Zeit und Raum**  
Visualisierung und Analyse von Geodaten über Transaktionen mit  
Basler Liegenschaften in der Vormoderne

- Patrick D. Brookshire und Jonas Richter  
 171 **Im Netz mittelhochdeutscher Wörter**  
 Das ‚Mittelhochdeutsche Wörterbuch‘ und seine Quellen
- Abdallah Mohamed Abdallah Abdelnaby, Nathanael Busch,  
 Julika Moos und Lilja Mareike Sautter  
 188 **Linked Open Bibliographical Data am Beispiel der  
 internationalen Artusforschung**  
 Wie Literaturangaben mehr leisten, als Literatur anzugeben
- Stefan Tomasek, Kiara Hart, Franziska Schulte und Christian Reul  
 206 **Automatische intratextuelle Kohärenzanalyse am Beispiel  
 des ‚Marienlebens‘ Philipps von Seitz**

## FORUM MITTELALTER

- 227 **Rezensionen**  
 Maryvonne Hagby: ‚Die Königstochter von Frankreich‘ (Monika Unzeitig) – Klaus Herbers u. Peter Rückert (Hgg.): Pilgern – Heil – Heilung (Christian Kaiser) – Christoph Kann u. Monika Steffens (Übers.): Heinrich von Gent (Julia Beier) – Fritz Peter Knapp (Übers.): La mort le roi Artu (Marion Darilek) – Jacqueline Murray (Hg.): Patriarchy, Honour, and Violence (Simon Suttman) – Sam Ottewill-Soulsby: The Emperor and the Elephant (Eric Böhme) – Hans-Joachim Schmidt: Neue Gesetze für ein besseres Leben? (Felix Grollmann) – Jörg Sonntag, Der spielende Gott des Mittelalters (Martin Thurner)



# Einleitung

## Schnittstelle Mediävistik.

### Kollaborationen der Mittelalterforschung im digitalen Zeitalter

#### Kontakt

**Dr. Luise Borek,**

Technische Universität Darmstadt,  
Institut für Sprach- und  
Literaturwissenschaft,  
Residenzschloss 1, 64283 Darmstadt,  
luise.borek@tu-darmstadt.de  
 <https://orcid.org/0000-0001-5849-374X>

**Hannah Busch,**

Universität zu Köln  
Cologne Center for eHumanities,  
Albertus-Magnus-Platz, 50923 Köln  
hannah.busch@uni-koeln.de  
 <https://orcid.org/0000-0002-3491-8202>

**Dr. Karoline Döring,**

Paris Lodron Universität Salzburg,  
Fachbereich Geschichte,  
Interdisziplinäres Zentrum für Mittel-  
alter und Frühneuzeit, Rudolfskai 42,  
5020 Salzburg, Österreich,  
karolinedominika.doering@plus.ac.at  
 <https://orcid.org/0000-0003-1943-1989>

**Dr. Nora Ketschik,**

Universität Stuttgart, Institut für  
Maschinelle Sprachverarbeitung,  
Pfaffenwaldring 5b, 70569 Stuttgart,  
nora.ketschik@ims.uni-stuttgart.de  
 <https://orcid.org/0000-0001-8758-5432>

**Dr. Katharina**

**Zepezauer-Wachauer,**  
Paris Lodron Universität Salzburg,  
Mittelhochdeutsche Begriffsdatenbank  
(MHDBDB), Fachbereich Germanistik,  
Interdisziplinäres Zentrum für  
Mittelalter und Frühneuzeit,  
Erzabt-Klotz-Straße 1,  
5020 Salzburg, Österreich,  
katharina.wachauer@plus.ac.at  
 <https://orcid.org/0000-0001-9310-9029>

**Abstract** The digital turn has profoundly reshaped medieval studies, positioning it as an interdisciplinary and collaborative field. The integration of digital tools and methodologies addresses challenges in preserving and analysing fragile medieval sources while enabling new scholarly approaches. This paper explores the dual importance of digital sustainability and thematic intersections in advancing the field. Digital sustainability ensures data and methods remain findable, accessible, interoperable, and reusable, in short, align with FAIR principles. Thematic intersections foster interdisciplinary dialogue by connecting distinct research domains and promoting innovative collaborations.

Using the tools and guidelines for text encoding from the Text Encoding Initiative (TEI), the Research Data Framework (RDF) as a method for semantic representation, and standards like the International Image Interoperability Framework (IIIF) for image interoperability, alongside approaches such as Linked Open Data (LOD), medievalists have built infrastructures that transcend disciplines. These tools facilitate detailed source analyses and contextualize medieval objects within broader cultural frameworks. Case studies from the ‘Network Linked Open Middle Ages’ show how collaborative infrastructures enhance sustainability and interconnectivity while fostering dialogue between humanities and technical domains.

**Keywords** Digital Medieval Studies; Digital Sustainability; Linked Open Data; Collaboration; Interoperability

## 1 Der Begriff der Schnittstelle

Die Mediävistik ist ein per definitionem interdisziplinär ausgerichtetes Fachgebiet. Aufgrund ihrer heterogenen, oftmals fragilen und nicht leicht zugänglichen Forschungsgegenstände haben sich Mediävist\*innen zu deren Erforschung schon früh der Möglichkeiten des Computers bedient und sich damit als Pionier\*innen der digitalen Geisteswissenschaften erwiesen. Bereits in der philologischen Tradition des Mittelalters wurden Werkzeuge und Praktiken entwickelt, die heute durch digitale Ansätze transformiert werden, wie etwa Übersetzungshilfen, Thesauri, Konkordanzen und Annotationen. Die Quellen adäquat digital zu repräsentieren und erfahrbar zu machen, stellt die Mediävistik vor Herausforderungen, denen sie mit der Nutzung digitaler Verfahren und Standards begegnet, etwa TEI (*Text Encoding Initiative*) für die digitale Textkodierung, IIIF (*International Image Interoperability Framework*) für die interoperable Bereitstellung von technischen Bilddaten oder das *Resource Description Framework* (RDF) für eine maschinenlesbare semantische Darstellung von Daten im Web.<sup>1</sup> So konnten in den letzten Jahren viele Ressourcen digital aufbereitet und nachhaltig zur Verfügung gestellt werden,<sup>2</sup> etwa im Bereich der Editionsphilologie für die Darstellung synoptischer Editionen oder die Einbindung von Digitalisaten. Auch die Bestrebungen der *Medieval Unicode Font Initiative* (MUFI) zur modernen Codierung der in mittelalterlichen Handschriften gebräuchlichen Abkürzungen, diakritischen Zeichen und Satzzeichen haben hier einen wesentlichen Beitrag geleistet. Die angewandten Standards sind mittlerweile etabliert und eröffnen immer umfangreichere Möglichkeiten, verschiedene Ressourcen miteinander in Bezug zu setzen, vorhandenes Wissen zu verdichten und neue Zugänge zu schaffen.

Dabei geht es längst nicht mehr nur darum, Quellen digital bereitzustellen. Vielmehr rückt in den Fokus, diese Quellen in den Gesamtzusammenhang ihrer Zeit sowie in den wissenschaftlichen Diskurs einzubetten.<sup>3</sup> Dennoch bleiben die bislang in Forschungsprojekten erarbeiteten digitalen Wissensplattformen häufig Insellösungen, die nur begrenzt mit weiteren Ressourcen interagieren. Nur in Ansätzen wurden bisher theoretische Überlegungen entwickelt, wie einzelne Projekte und ihre Ergebnisse systematisch miteinander verknüpft werden können. Zugleich ist evident, dass das Ziel der digitalen Erschließung ein verdichtetes Wissensnetz sein muss, das nicht nur einzelne Gegenstände miteinander verbindet,

---

1 Diese und andere Abkürzungen und Fachbegriffe im Themenheft sind erklärt in Klug 2024.

2 Vgl. die Ressourcensammlung bei van Haaren u. Diekkämper 2024. Auch im Kontext des ‚Netzwerkes Offenes Mittelalter‘ ist eine Wissensplattform in Vorbereitung, die Ressourcen, Methoden und Tools sammelt und online zugänglich macht.

3 Siehe Burrichter, Gebert, Mackert u. Viehhauser 2021.

sondern auch die gesamte Überlieferung in den Blick nimmt – eine Überlieferung, die per se fragmentarisch und zudem nicht in Gänze erschlossen ist.

Zugleich wächst das Bewusstsein dafür, historische Gegenstände so aufzubereiten, dass sie über ihre ursprünglichen Kontexte hinaus anschlussfähig bleiben und für neue, noch unbekannte Forschungsfragen genutzt werden können. Diese Aufgaben stellen besondere Herausforderungen dar, denn sie berühren die Historizität und die kulturhistorischen Bezugssysteme der Forschungsgegenstände. Entsprechend handelt es sich um hochgradig interdisziplinäre Prozesse, die eine enge Zusammenarbeit innerhalb der gesamten mediävistischen Community erfordern.

Vor diesem Hintergrund zeichnet das Themenheft ein Bild aktueller mediävistischer Kollaborationen und gewährt Einsichten in die Mehrwerte, die entstehen, wenn qualitative und vernetzte Forschungsdaten die Grundlage für innovative Forschungsfragen bilden. Der Titel ‚Schnittstelle Mediävistik – Kollaborationen der Mittelalterforschung im digitalen Zeitalter‘ verweist bewusst auf die Vielschichtigkeit und Polyvalenz des Begriffs der ‚Schnittstelle‘. Er steht nicht nur für die interdisziplinäre Schnittstelle, die als Verbindungsglied zwischen verschiedenen Fachrichtungen fungiert, sondern auch für die mediale Schnittstelle, die einerseits Interoperabilität zwischen verschiedenen digitalen Angeboten ermöglicht, andererseits aber auch analoge und digitale Formen der Quellenrepräsentation erlaubt.

## 2 Genese des Themenhefts: Impulse aus der mediävistischen Community

Einen zentralen Impuls zu dieser Idee eines verdichteten Wissensnetzes lieferte das DFG-geförderte ‚Netzwerk Offenes Mittelalter‘, an dem die Autorinnen der Einleitung mitwirkten.<sup>4</sup> Das Netzwerk explorierte innovative Verfahren zur weiteren Vernetzung der unterschiedlichen Datenbestände im Kontext interdisziplinärer Forschung. Ein besonderes Augenmerk wurde dabei auf *Linked Open Data* (LOD) gelegt. Darunter versteht man miteinander verknüpfte, offen zugängliche Daten, die Ressourcen miteinander verbinden und hierfür auf bestimmte Technologien (wie die Auszeichnungs- und Abfragesprachen RDF, OWL und SPARQL) zurückgreifen. Diese Verknüpfung strukturierter und eindeutig identifizierter Entitäten optimiert nicht nur die Qualität und Erschließungstiefe von Daten, sondern ermöglicht auch ihre Kontextualisierung und damit ein tieferes Verständnis der strukturellen und semantischen Beziehungen zwischen Forschungsdaten. Gleichzeitig stellt sie die Interoperabilität digital-geisteswissenschaftlicher Projekte sicher.

---

4 Zur Arbeit des Netzwerks siehe <https://offenesmittelalter.org> (Zugriff: 19.03.2025).

Das Netzwerk erprobte und entwickelte anhand konkreter Ressourcen und Forschungskontexte digitale Lösungen für die Mediävistik, die die Qualität der Daten wie deren Erforschung verbessern und zu einem integralen Bestandteil des mediävistischen Methodeninventars werden können. Denn der Einsatz von LOD-Verfahren stellt eine vielversprechende Option dar, Ressourcen zu erschließen, weiter anzureichern und somit sichtbar, verfügbar und nachnutzbar zu machen. Gleichzeitig werden sie den FAIR-Prinzipien (*Findable, Accessible, Interoperable, and Reusable*), nach denen Forschungsdaten auffindbar, zugänglich, interoperabel und nachnutzbar sein sollen, gerecht und spiegeln damit den Anspruch des interdisziplinären, offenen Netzwerks im Sinne von Open Science wider.

Unter zeitgemäßer digitaler Mittelalterforschung verstehen wir auch nationale und internationale Forschungsinfrastrukturen, die von Mediävist\*innen mit einem gemeinsamen Interesse an einer tiefgreifenden Erschließung von Forschungsgegenständen interdisziplinär, multimodal und international genutzt sowie kollaborativ gepflegt und aktiv mitgestaltet werden. Die im Namen des Netzwerks verankerte Offenheit bezieht sich dabei nicht nur auf die Verwendung offener Standards und die interdisziplinären Perspektiven auf heterogene Forschungsgegenstände, sondern auch auf eine gesamtgesellschaftliche Teilhabe am zu bewahrenden Kulturerbe. Eine fundierte Erschließung dieses Kulturerbes und ein optimierter Zugriff auf den entstehenden gemeinsamen Wissensfundus sind dafür unverzichtbare Voraussetzungen.

Methodisch zeigten *Linked Data*-Verfahren im Rahmen der Netzwerkaktivitäten exemplarisch, dass digitale Verfahren an bestehende geisteswissenschaftliche Praktiken anknüpfen. Hierzu zählen etwa das Identifizieren, Disambiguieren und Kontextualisieren – Aktivitäten, die zur Kulturpraxis des Annotierens gehören. Jede Auszeichnung von Daten sowie jede Form des Markups hat eine mehr oder weniger ausgeprägte Kommentarfunktion. Digitale Verfahren haben gegenüber analogen den Vorteil, dass Auszeichnungen, wenn sie einem Standard folgen, von der mediävistischen Community nicht nur nachgenutzt, sondern auch sukzessive angereichert werden können.<sup>5</sup> Die Enkodierung fungiert dabei als Instrument zur Dekodierung enthaltenen Wissens – und trägt in allen mediävistisch arbeitenden Fächern zum Erkenntnisgewinn bei.

---

5 Vgl. die verschiedenen Ergebnisse der Netzwerkarbeit: Borek, Zeppezauer-Wachauer u. Ketschik 2022; Ketschik u. Grüntgens 2024; Zeppezauer-Wachauer u. Heiles 2023; Busch u. Müller 2023.

### **3 Interdisziplinarität: Vernetztes Denken als Grundlage einer integrativen, digitalen Mediävistik**

Die Arbeit des Netzwerks findet unter anderem ihren Niederschlag in diesem Themenheft. Es nimmt Impulse auf, die aus dem intensiven Austausch von insgesamt 38 digital arbeitenden Mitgliedern und Assoziierten aus verschiedenen Sprach- und Literaturwissenschaften, aus den Geschichtswissenschaften und aus der Kunstgeschichte zu den gemeinsamen Themen ‚Text‘ (2021), ‚Daten‘ (2022), ‚Objekte‘ (2022), ‚Methodenkritik‘ (2023) und ‚Schnittstelle Mediävistik‘ (2023) gewonnen wurden. Darüber hinaus brachten auch Expert\*innen unterschiedlichster Disziplinen, die nicht dem Netzwerk angehörten, wertvolle Perspektiven und Anregungen ein. Diese Kollaborationen sind Nachweis genug, dass die digitale Mediävistik von ihrer inhärenten Interdisziplinarität lebt, die weit über die Verbindung einzelner Fachgebiete hinausgeht. Das Themenheft zeigt exemplarisch, wie Interdisziplinarität auf zwei Ebenen wirksam wird: einerseits auf der Ebene der Forschungsgegenstände, die von historischen Daten und literarischen Texten bis hin zu Schrift-Bild-Bezügen reichen; andererseits auf der Ebene der Methodenvielfalt, die sowohl ‚traditionelle‘ Arbeitsweisen wie das Edieren und Analysieren von Texten als auch computationale Ansätze wie Annotation, Modellierung, maschinelles Lernen und LOD umfasst.

Diese Methodenvielfalt ist kein Selbstzweck, sondern ein produktiver Zugang zur Reflexion über die Grenzen und Potentiale der digitalen Mediävistik. Der Zwang zur Formalisierung, der etwa durch die Modellierung historischer Daten oder die Anwendung von Ontologien entsteht, generiert neue Fragestellungen, die ohne die digitale Perspektive nicht denkbar wären. So wird etwa die Normalisierung von Datenformaten oder die Integration heterogener Forschungsgegenstände selbst zum Untersuchungsgegenstand. Interdisziplinarität in der digitalen Mediävistik bedeutet unserem Verständnis nach mehr als nur Multidisziplinarität, also die parallele Anwendung verschiedener Methoden und die gleichzeitige Bearbeitung von Forschungsthemen durch verschiedene Disziplinen. Sie bedeutet für uns auch Multiperspektivität, also die Vielzahl der Sichtweisen auf kollaborativ bearbeitete Forschungsthemen und Forschungsgegenstände, und sie meint besonders eine dynamische Verschmelzung und Weiterentwicklung der Ansätze zu neuen, eigenständigen Fragestellungen in dezidiert vielstimmigen Forschungsperspektiven.

Die Beiträge des Themenhefts greifen diese Dynamik auf und zeigen, wie digitale Werkzeuge disziplinübergreifende Verbindungen schaffen und dadurch bestehende Strukturen kritisch hinterfragen. Die darin vorgestellten Ansätze verdeutlichen, dass die digitale Mediävistik nicht als Hilfsdisziplin agiert, sondern ein eigenständiges Paradigma darstellt, das die Vielfalt der Gegenstände und

Methoden produktiv miteinander verknüpft und sie in einem von vornherein reziprok angelegten Forschungsprozess der beteiligten Expert\*innen bearbeitet. Die bereichernden Effekte einer so verstandenen Interdisziplinarität von mediävistischen Kollaborationen im digitalen Zeitalter macht das Themenheft sichtbar, indem es sich der computationell arbeitenden Mittelalterforschung in ihrer verbindungsstiftenden Funktion widmet.

#### 4 *Digital turns*: Forschungsprozesse und Methoden im Wandel

Über fünf Jahre sind vergangen, seitdem in dieser Zeitschrift das Themenheft mit dem Schwerpunkt ‚Digitale Mediävistik‘ erschien.<sup>6</sup> Im Hinblick auf die rasanten Entwicklungen in den textverarbeitenden und textgenerierenden Disziplinen scheint dies eine lange Zeit zu sein. Eine Zeit, in der sich ein einschneidender *computational turn* vollzogen hat, nämlich von computergestützten zu zunehmend automatisierten Verfahren wie zum Beispiel dem *Deep Machine Learning*, unter das auch als ‚Künstliche Intelligenz‘ (KI) bekannte Anwendungen fallen. Gerade durch – besonders für Laien – undurchsichtige KI-Verfahren scheint sich die Kluft zwischen Expert\*innen in den Bereichen der Computerphilologie und den traditionell ausgebildeten Mediävist\*innen erneut zu verstärken. Dabei nimmt die digitale Mediävistik bereits seit ihren Anfangstagen eine wichtige Position in der Brückenbildung zwischen der Informatik und den mediävistischen Disziplinen ein.

Für einen Rückblick auf die Geschichte der digitalen Mediävistik sowie ihre Rolle bei der Etablierung der digitalen Geisteswissenschaften nicht nur im deutschsprachigen Raum verweisen wir an dieser Stelle auf die Einleitung im einschlägigen Heft von 2019.<sup>7</sup> Die Herausgeber betonen dort den Perspektivwechsel und den Wandel von Fragestellungen als Folge der Anwendung von digitalen Hilfsmitteln und Methoden ebenso wie die Erleichterung, Beschleunigung und Verbesserung von Prozessen und Analyseverfahren. Andrea RAPP fasste im Rahmen ihrer Keynote beim vierten Arbeitstreffen des DFG-Netzwerks ‚Netzwerk Offenes Mittelalter‘ in Salzburg,<sup>8</sup> das wie auch dieses Themenheft der ‚Schnittstelle Mediävistik‘ gewidmet war, die Geschichte der germanistischen Mediävistik zusammen und wies auf die im Epochenbegriff enthaltene Funktion hin: „Ein mittleres Zeitalter muss eine Schnittstelle sein.“<sup>9</sup>

---

6 Vgl. Bleier, Fischer, Hiltmann u. a. 2019.

7 Vgl. ebd.

8 Verwiesen sei an dieser Stelle auf den Bericht zum Arbeitstreffen auf dem Mittelalterblog: van Beek u. Borek 2023.

9 Rapp 2023. Der Mitschnitt des Vortrags wurde von uniTV, dem Salzburger Unifernsehen, aufgezeichnet und zur Nachschau veröffentlicht.

In der Zeitschrift ‚Digital Humanities Quarterly‘ befasst sich Edgar LEJEUNE in seinem Artikel „‚Open‘ or ‚Close‘ Research Instruments? Conflicting Rationales in the Organization of Early Digital Medieval History in Europe (1960–1990)“ mit der verbindungsstiftenden Funktion der frühen digitalen Mediävistik.<sup>10</sup> Der Fokus des Artikels liegt weniger auf Projekten, sondern auf den Debatten, die zu den ersten computerbasierten Entwicklungen geführt haben. LEJEUNE kommt zu dem Ergebnis, dass die digitale Mediävistik schon immer von Kollaborationen geprägt war, und zwar von Kollaborationen auf verschiedenen Ebenen. Zum einen konnten und können neue Tools und Methoden nur als Kollektiv, also in enger Zusammenarbeit zwischen Entwickler\*innen und Fachwissenschaftler\*innen entwickelt werden. Zum anderen waren die einzelnen Kollektive schon früh untereinander vernetzt und im Austausch darüber, wie Forschungsinstrumente und -daten auch von anderen Forschenden genutzt werden können.<sup>11</sup> Somit ist die digitale Mediävistik bereits seit den siebziger Jahren bestrebt, offene und nachnutzbare Lösungen zu schaffen und Insellösungen zu vermeiden. Die Geschichte der digitalen Mediävistik sei daher stets auch die Geschichte von interdisziplinären Kollaborationen über die Schnittstelle der geisteswissenschaftlichen Fächer hinaus gewesen.<sup>12</sup> Um die Entwicklung der digitalen Mediävistik in diesem Beitrag gewissermaßen fortzuschreiben, sei auf einige wegweisende Publikationen der letzten Jahre verwiesen. Mit dem Band ‚Digital Humanities in den Geschichtswissenschaften‘ von Christina ANTENHOFER, Christoph KÜHBERGER und Arno STROHMEYER<sup>13</sup> liegt ein interdisziplinäres Studienbuch vor, das – nicht nur für Historiker\*innen – relevante Bereiche der digitalen Geisteswissenschaften abdeckt.

Im Frühjahr 2022 erschien in der Zeitschrift ‚Science‘ der Artikel ‚Forgotten Books: The Application of Unseen Species Models to the Survival of Culture‘<sup>14</sup>, der als ein Musterbeispiel für multidisziplinäre Kollaboration gesehen werden kann. Der Artikel ist das Ergebnis einer Zusammenarbeit zwischen Expert\*innen in den digitalen Geisteswissenschaften, der Mediävistik sowie den Umweltwissenschaften, die ein statistisches Modell zur Berechnung der Größe und Diversität von Tierpopulationen verwenden, um die Überlieferungs- und Verlustquotienten zu ermitteln. Durch die Veröffentlichung in der Zeitschrift ‚Science‘ erlangte dieser Beitrag Bekanntheit und Bedeutung über die Geisteswissenschaften hinaus.

Neben dem *computational turn* in Richtung automatisierter Verfahren der Erfassung, Analyse und Auswertung ist auch die verfügbare Datenmenge in den

---

10 Lejeune 2024.

11 Vgl. ebd. sowie Genicot 1975.

12 Vgl. ebd.

13 Antenhofer, Kühberger u. Strohmeyer 2023.

14 Kestemont, Karsdorp, de Bruijn u. a. 2022.

letzten Jahren deutlich angestiegen. Mit der breiten Verfügbarkeit generativer und automatisierter Verfahren gewinnt auch die Anforderung an standardisierte, offene und dadurch nachhaltigere Daten eine neue Dringlichkeit. Denn nur durch nachhaltig verfügbare und genormte Daten, durch die Einhaltung von Standards und die Verwendung persistenter Identifikatoren kann eine langfristige interdisziplinäre Forschung bzw. Nachnutzung der Daten im Wechselspiel verschiedener Disziplinen und Methoden ermöglicht werden.

In diesem Sinne tragen nationale und internationale Forschungsinfrastrukturen, die von Mediävist\*innen mit einem gemeinsamen Interesse an einer tiefgreifenden Erschließung von Forschungsgegenständen kollaborativ gepflegt und gestaltet werden, zu einer zeitgemäßen digitalen Mittelalterforschung bei. Unterstützung kommt hier durch die Einrichtung der geisteswissenschaftlichen NFDI-Konsortien, die eine fächerübergreifende Anlaufstelle für Services und Beratung sind.<sup>15</sup>

Darüber hinaus sind viele Herausforderungen nach wie vor aktuell und bekommen durch die sich rasant entwickelnden Technologien eine neue Relevanz. Diese Herausforderungen spiegeln sich auch in den Beiträgen des Themenhefts wider. Zur Entwicklung neuer Methoden, bei der die Anwendung neuer Technologien erprobt werden, braucht es ausreichend Trainingsdaten, die häufig noch nicht in der benötigten Form vorhanden oder zugänglich sind. Nicht-normierte Daten erschweren den Einsatz von KI, beispielsweise beim *Word-Embedding*, so dass gegebenenfalls Fehler wieder und wieder reproduziert werden. Das betrifft vor allem frühe, nicht-normierte Sprachstufen sowie Sprachen, für die nur geringe Sprachdaten und sehr kleine linguistische Korpora vorliegen, sogenannte *Low Resource Languages*. Der Beitrag von Stefan TOMASEK, Kiara HART, Franziska SCHULTE und Christian REUL führt zum Beispiel an einer Stelle aus, dass die eigentlich naheliegende Analyse über ein *Word-Embedding*-Verfahren aufgrund der mangelhaften Modelle nicht brauchbar war. Hier wäre die mediävistische Community gefragt, ihre Daten für Training und Weiterentwicklung geeigneter

---

15 Seit der Einrichtung der Nationalen Forschungsdateninfrastruktur (NFDI) 2020 sind vier für die Mediävistik relevante Konsortien in die Förderung aufgenommen worden: Text+ für „sprach- und textbasierte Forschungsdaten“, NFDI4Culture für „Forschungsdaten zu materiellen und immateriellen Kulturgütern“, NFDI4Objects für „die materiellen Hinterlassenschaften der Menschheitsgeschichte“ sowie NFDI4Memory für die „historisch arbeitenden Geisteswissenschaften“ (siehe <https://www.nfdi.de/konsortien/> [Zugriff: 15.12.2024]). Zentrale Aufgaben der Konsortien liegen in der Integration der Forschungscommunities, der Publikation und Sammlung von Standards sowie der Vergabe von kleinen Stipendien zur Unterstützung von Forschenden. Text+ und NFDI4Culture bieten darüber hinaus bereits weitere Services an, wie die Beratung von Forschenden und das Angebot von Repositorien für (Meta-)Daten, vgl. <https://nfdi4culture.de/resources.html> (Zugriff: 15.12.2024) und <https://text-plus.org/> (Zugriff: 15.12.2024). Die NFDI-Konsortien sind untereinander vernetzt und bieten auch einen übergreifenden Austausch an.

Modelle zur Verfügung zu stellen und durch gemeinsame Anstrengungen einen integren und nachhaltigen Einsatz der neuen Technologien zu gewährleisten.

## **5 Community und Kollaboration: Nachhaltigkeit und Netzwerke als Basis der digitalen Mediävistik**

Die digitale Mediävistik basiert auf dynamischen Netzwerken, die Kollaboration und nachhaltige Praktiken miteinander verbinden. Diese Zusammenarbeit geht weit über punktuelle Kooperationen hinaus: Sie zielt darauf ab, nachhaltige Netzwerke zu schaffen, in denen Wissen, Methoden und Daten zwischen verschiedenen Disziplinen und Institutionen ausgetauscht und weiterentwickelt werden können. Ein zentraler Aspekt dabei ist der Verknüpfungsgedanke, der die Integration verschiedener Ansätze ermöglicht sowie neue Synergien schafft.

Das Themenheft selbst ist ein Beispiel für die produktive Kraft kollaborativer Netzwerke. Die Beiträge setzen auf gemeinsame Standards, disziplinübergreifende Ansätze und die Entwicklung nachhaltiger Forschungsdaten. Michael SCHONHARDT beispielsweise zeigt anhand digitaler Editionen, wie essentiell es ist, diese als dynamische Schnittstellen zu begreifen, damit sie nicht nur dem aktuellen Forschungsbedarf gerecht werden, sondern sich auch an zukünftige technologische Entwicklungen anpassen. Er betont, dass die Nachnutzbarkeit und Flexibilität der Datenformate zentral sind, damit diese langfristig relevant bleiben. Sabine TITTEL legt den Fokus auf die Anwendung von historisierten Ontologien in der linguistischen Forschung. Sie zeigt, wie LOD und Ontologien genutzt werden können, um historische Sprachdaten interoperabel zu machen und semantisch zu verknüpfen. Dies ermöglicht eine bedeutungsbasierte Recherche über verschiedene Sprachstufen hinweg und schafft Verbindungen zu kulturell-historischen Kontexten, die über die Grenzen einzelner Datensilos hinausgehen. Abdallah Mohamed Abdallah ABDELNABY, Nathanael BUSCH, Julika MOOS und Lilja Mareike SAUTTER illustrieren die Potentiale von LOD am Beispiel der internationalen Artusforschung. Ihr Beitrag zeigt, wie bibliographische Daten als verknüpfte Forschungsressourcen genutzt werden können, die nicht nur bestehendes Wissen sichern, sondern durch Verknüpfung mit anderen Datenbanken neue interdisziplinäre Fragestellungen ermöglichen. Dabei wird deutlich, wie wichtig ein nachhaltiger Umgang mit Daten und eine enge Zusammenarbeit zwischen Forschenden und Infrastrukturen für die langfristige Nutzbarkeit solcher Projekte sind. Bernhard BAUER, Roman BLEIER und Jakob SONNBERGER gehen auf die Rolle von LOD in der Beschreibung mittelalterlicher Handschriften ein. Sie illustrieren, wie durch die Nutzung standardisierter Auszeichnungssprachen wie TEI und RDF sowie durch den Einsatz von Technologien wie IIF die Interoperabilität und Nachnutzbarkeit von Daten in der Mediävistik verbessert werden kann. Ihr

Beitrag hebt hervor, dass diese Technologien die Vernetzung der Daten fördern und dadurch eine fundierte Grundlage für die digitale Nachhaltigkeit schaffen. Diese Ansätze unterstreichen die Bedeutung kollaborativer Verantwortung innerhalb der Community und zeigen, wie die digitale Mediävistik nicht nur bestehende Ressourcen effektiv nutzt, sondern auch nachhaltige Infrastrukturen entwickelt, um zukünftige Forschungsbedarfe zu antizipieren und die Community langfristig zu unterstützen.

## **6 Von Brücken zu Netzwerken: Kollaborative Verantwortung in der Community**

Die integrative, digitale Mediävistik, die mit dem Begriff der Schnittstelle adressiert wird, versteht sich nicht mehr nur als Brücke zwischen festen Punkten, etwa den ‚analogen‘ Fachdisziplinen und ‚der IT‘. Sie ist vielmehr ein dynamisches Netzwerk geworden, das Verbindungen in alle Richtungen schafft. Dieses Netzwerk ist nicht statisch, sondern entwickelt sich durch die kontinuierliche Zusammenarbeit innerhalb der Community weiter. Durch den Einsatz modularer Werkzeuge und die Nutzung offener Plattformen entstehen Strukturen, die die aktuelle Forschung fördern und auch für zukünftige Entwicklungen eine geeignete Basis bilden. Ein solches Netzwerk erfordert nicht nur technologische Lösungen, sondern auch eine klar formulierte Verantwortung der Community. Diese Verantwortung zeigt sich beispielsweise in der Nutzung und Weiterentwicklung bestehender Ressourcen, wie sie durch nationale und internationale Forschungsdateninfrastrukturen wie CLARIAH-DE/-AT oder die Konsortien der NFDI (hier insbesondere Text+, NFDI4Culture, NFDI4Memory und NFDI4Objects) gefördert werden. Derartige Verbünde zielen darauf ab, die Langzeitarchivierung und Nachnutzbarkeit von Forschungsdaten zu gewährleisten und dabei modulare und flexible Werkzeuge zu entwickeln, die sich an neue Anforderungen anpassen lassen. Digitale Nachhaltigkeit bedeutet dabei mehr als reine Datensicherung: Sie umfasst auch die Sicherstellung methodischer Qualität und die langfristige Anschlussfähigkeit wissenschaftlicher Erkenntnisse an neue Forschungskontexte. Nur durch die konsequente Einhaltung dieser Prinzipien kann eine digitale Forschungslandschaft entstehen, die zukunftsfähig bleibt.<sup>16</sup> Diese Perspektive unterstreicht die Bedeutung eines reflektierten Umgangs mit digitalen Ressourcen und zeigt, wie kollaborative Verantwortung und nachhaltige Strategien die digitale Mediävistik langfristig stärken können.

---

<sup>16</sup> Vgl. Zeppezauer-Wachauer, Zangerl u. Stigler 2024.

Kollaboration in der digitalen Mediävistik bedeutet auch, Verantwortung für die Nutzung und Weiterentwicklung bestehender Ressourcen zu übernehmen. Die ‚Digital Humanities Austria Strategie 2021+‘<sup>17</sup> des CLARIAH-AT-Konsortiums betont etwa die Bedeutung eines harmonisierten Repositorienverbunds für geisteswissenschaftliche Daten, um Langzeitarchivierung und Nachnutzbarkeit zu sichern. Gleichzeitig müssen flexible Standards geschaffen werden, die sich an neue technologische und wissenschaftliche Entwicklungen anpassen lassen. Ein weiterer zentraler Aspekt ist die Berücksichtigung von Fragen des offenen Zugangs und der Rechteverwaltung. Open-Science-Prinzipien wie Transparenz und Nachnutzbarkeit schaffen hier eine Grundlage, stoßen jedoch auf rechtliche und technische Herausforderungen. Überlegungen zur Zugänglichkeit von Daten sowohl von juristischer als auch von technischer Seite betreffen die gesamte Community. Die Diskussion dieser Fragen trägt dazu bei, die digitale Mediävistik als offenes und inklusives Forschungsfeld zu gestalten.

## 7 Die Schwerpunkte des Themenhefts

Das Themenheft zeugt von der kollaborativen Arbeitspraxis der digital arbeitenden Mittelalterforschung sowie von einer im Sinne digitaler Nachhaltigkeit verstandenen Vernetzung der Datenbestände. Die Beiträge diskutieren und evaluieren anhand konkreter Praxisbeispiele die hierbei angewandten Methoden und Praktiken. Dabei verstehen sie sich als Momentaufnahmen von im Wandel begriffenen Disziplinen, die durch methodische Offenheit zum einen interoperabel und zum anderen anschlussfähig für künftige Entwicklungen bleiben.

Die Inhalte des Themenhefts reichen von methodologischen über daten- oder objektbezogene Fragestellungen bis hin zu infrastrukturellen Aufgaben, die unter der Perspektive der digitalen Nachhaltigkeit und Vernetzung vereint werden.

Eine zentrale Herausforderung ist nach wie vor die Modellierung von Informationen, insbesondere hinsichtlich spezifischer Anforderungen, die sich aus der Arbeit mit historischem Material ergeben. Hier kann die Mittelalterforschung nicht immer auf allgemeine Standards zurückgreifen, sondern muss (auch) fachspezifische Lösungen hervorbringen. Dieser Problematik widmet sich unter anderem TITTEL, die in ihrem Beitrag aufzeigt, wie historisierte Ontologien Veränderungen von Begriffsbedeutungen und Konzepten modellieren können und damit einen historisch adäquaten Zugriff ermöglichen. Sie schlägt spezifische Ontologien für Bereiche wie mittelalterliche Medizin und Recht vor, um anachronistische Zuordnungen zu vermeiden und historisch adäquate Datenabfragen zu gewährleisten. In

---

<sup>17</sup> Vgl. CLARIAH-AT Konsortium 2021.

Kombination mit LOD wird auf diese Weise eine präzise Modellierung historischer Begriffsveränderungen ermöglicht.

Auch weitere Beiträge des Themenhefts thematisieren die Notwendigkeit fachspezifischer Lösungsansätze in der Mittelalterforschung. Elisabeth GRUBER und Peter FÄRBERBÖCK gehen beispielsweise auf die Herausforderung der Heterogenität von Quellen ein und stellen das Tool ‚ITEM‘ vor. Dieses wurde entwickelt, um pragmatische Schriftquellen wie Inventare und Rechnungsbücher, die systematisch annotiert und in ein graphbasiertes Datenmodell überführt worden sind, abzufragen. Sie zeigen, dass hier nicht nur kontrollierte Vokabulare, sondern auch spezifisch an die Quellenstruktur angepasste Ontologien notwendig sind, um eine nachhaltige Integration in das *Semantic Web* zu schaffen.

LOD als Verfahren der Verknüpfung und Kontextualisierung gewährleistet nicht nur eine nachhaltige und interoperable Modellierung von Daten, das Verfahren kann auch gewinnbringend mit anderen Ansätzen zusammengeführt werden. Neben TITTEL verbinden auch BAUER, BLEIER und SONNBERGER den kontextualisierenden LOD-Gedanken mit einem historisierenden ontologischen Ansatz. Sie untersuchen, wie durch die Kombination von Auszeichnungssprachen wie TEI und RDF einerseits und Bildtechnologien wie IIIF andererseits die Interoperabilität und Nachnutzbarkeit von Editionen erheblich verbessert werden kann. Sie heben hervor, dass offene Standards und flexible Datenmodelle entscheidend sind, um die Anforderungen der Forschung an mittelalterliche Handschriften langfristig zu erfüllen. Dabei betonen sie aber auch, dass die Interoperabilität von Metadaten nur durch speziell entwickelte Ontologien und angepasste Datenmodelle erreicht werden kann, da allgemeine Standards wie TEI zwar eine Grundlage bilden, aber nicht alle Anforderungen mittelalterlicher Manuskriptforschung abdecken können.

Auch SCHONHARDT verbindet am Praxisbeispiel der digitalen Edition des *Decretum Burchardi* editionsphilologische Ansätze mit kontextualisierenden und nachhaltigen Verfahren. Er konzentriert sich in seinem Beitrag auf den interdisziplinären Schnittstellengedanken der digitalen Mediävistik und plädiert dafür, digitale Editionen nicht isoliert zu entwickeln. Stattdessen befürwortet er offene Infrastrukturen, die eine hohe Wiederverwendbarkeit ermöglichen und nahtlos in größere Netzwerke integriert werden können. Sein Beitrag unterstreicht, dass methodische Offenheit und nachhaltige Strukturen essentielle Grundlagen für eine zukunftsorientierte Editionsphilologie sind.

Dana EICHHORST, Laura FRANK, Danah TONNE und Katharina H. WENDL betonen ebenfalls, dass klassische (analoge) Editionsmethoden nicht ausreichen, um die Varianten und intertextuellen Bezüge komplexer Textkorpora wie jüdischer Kommentarliteratur zu erfassen. Digitale Werkzeuge wie *Web Annotation Data Models* ermöglichen eine mehrdimensionale, a-lineare Edition, die historische Verbindungen zwischen Texten sichtbar macht und den Zugang zu bisher schwer zugänglicher Überlieferung verbessert.

Ansätze dieser Art machen deutlich, dass der Einsatz von LOD weit über die bloße Organisation von Daten hinausgeht. Er bietet die Möglichkeit, historische Zusammenhänge durch semantische Verknüpfungen transparent zu machen und Daten in ihrer spezifischen Bedeutung zu erschließen. Die Verbindung von LOD mit spezialisierten Ontologien ermöglicht zum einen eine historisch adäquate Modellierung und kann zum anderen auch zukünftige Entwicklungen antizipieren. Für die Editionsphilologie stellen quantitative und kontextualisierende Verfahren ein Werkzeug dar, das ebenso zur nachhaltigen Nutzung digitaler Editionen wie zur weiteren Generierung von Wissen führen kann.

ABDELNABY, BUSCH, MOOS und SAUTTER greifen diesen Gedanken auf und verdeutlichen in ihrem Beitrag zur Artusforschung, wie LOD-Verfahren bibliographische Daten nicht nur strukturieren, sondern durch semantische Kontextualisierung bereichern. Sie zeigen, dass die flexible Verknüpfung von Datenbanken es ermöglicht, vielfältige Perspektiven zu eröffnen und innovative Fragestellungen zu entwickeln. Die durch LOD geschaffene Interoperabilität wird hier als Schlüssel betrachtet, um Datenbestände nachhaltig zu nutzen und gleichzeitig den interdisziplinären Austausch zu fördern.

Neben der Erschließung und Analyse textbasierter Quellen eröffnet die digitale Mediävistik auch neue Möglichkeiten für die Untersuchung nicht-textueller Materialien. Besonders die Bereiche Realienkunde und Kunstgeschichte profitieren von innovativen digitalen Ansätzen, die Bilder und Objekte in ihren historischen Kontexten analysieren.

GRUBER und FÄRBERBÖCK verdeutlichen, wie pragmatische Schriftquellen wie Inventare und Rechnungsbücher genutzt werden können, um historische Objekte und ihre Kontexte digital zu erforschen. Sie zeigen, dass diese Verbindung textlicher und materieller Informationen neue Einblicke in den historischen Gebrauch und die Bedeutung von Objekten ermöglicht. Während bei Texten die Transkription oft den Ausgangspunkt für Analyse und Retrieval bildet, ist bei Bildquellen die semantische Annotation, wie Isabella NICKA, Miriam LANDKAMMER, Andreas UHL, Michael LINORTNER und Johannes SCHUIKI ausführen, eine wichtige Basis für disziplinäre und interdisziplinäre Forschung. Der Beitrag verweist auf die Notwendigkeit, geeignete Ontologien und Datenmodelle für Bildinhalte zu entwickeln, und evaluiert die Qualität von Bildannotationen unterschiedlicher Ersteller\*innen (Expert\*innen, Computer-Algorithmen und *Citizen Scientists*). Im Fokus stehen Annotationen visueller Eigenschaften von Bildelementen (zum Beispiel Holzmaserungen) und die Probleme bei ihrer Erstellung, aber auch ihr Mehrwert für die vergleichende Analyse visueller Strategien (*Distant Viewing*).

Die Betrachtung der semantischen Annotation von Bildquellen macht deutlich, wie wichtig standardisierte Methoden und Modelle für die weitere interdisziplinäre Forschung sind. Ähnlich wie bei Bildquellen spielt auch bei textbasierten Projekten wie dem ‚Mittelhochdeutschen Wörterbuch‘ (MWB) die Verknüpfung

von Daten und die Entwicklung effizienter Schnittstellen eine zentrale Rolle. Dabei steht unter anderem der Nachnutzungsgedanke im Vordergrund, wie Patrick D. BROOKSHIRE und Jonas RICHTER in ihrem Beitrag hervorheben. Die Nachnutzung wird nicht nur durch Textkodierungsstandards (TEI), sondern insbesondere durch Schnittstellen wie die föderierte Architektur des NFDI-Konsortiums Text+ ermöglicht, über die die Daten des MWB auch jenseits der Projektgrenzen abgefragt werden können. Dieser Fokus auf die Nachnutzung wird in anderen Beiträgen des Themenhefts aus interdisziplinärer Perspektive weitergeführt.

Die interdisziplinäre Nachnutzung von Geodaten steht im Zentrum des Beitrags von Benjamin HRTZ und Tobias HODEL, die am Beispiel der Basler Stadtgeschichte von 1300 bis 1700 zeigen, wie geographische Informationssysteme (GIS) und digitale Visualisierungswerkzeuge eingesetzt werden können, um Grundbesitz und Besitzwechsel analysierbar zu machen. Ihre Methodik verbindet räumliche Daten nicht nur mit historischen Fragestellungen, sondern integriert die Visualisierung als zentralen Bestandteil der Methodenentwicklung. Dadurch werden komplexe Zusammenhänge nicht nur sichtbar gemacht, sondern auch für weitere Forschungsfragen zur Wirtschaftsgeschichte des Mittelalters nachhaltig nutzbar. Besonders betont wird, wie modulare Schnittstellen und interoperable Standards die Nachnutzung dieser Daten über disziplinäre und projektbezogene Grenzen hinweg ermöglichen.

Neben der Nachnutzung von Daten spielen Normdaten eine zentrale Rolle bei der Sicherstellung von Interoperabilität und nachhaltiger Forschung. Die Bedeutung von Normdaten wie der ‚Gemeinsamen Normdatei‘ (GND) wird unter anderem bei ABDELNABY, BUSCH, MOOS und SAUTTER hervorgehoben. Ihr Beitrag zeigt, wie die GND genutzt werden kann, um Forschungsdaten miteinander zu verknüpfen und dadurch eine interoperable Nutzung zu ermöglichen. Sie demonstrieren, dass Normdaten nicht nur für die Standardisierung, sondern auch für die semantische Kontextualisierung von Informationen eine Schlüsselrolle spielen. Auch in anderen Beiträgen, etwa von BAUER, BLEIER und SONNBERGER, wird die GND als essentielles Werkzeug für die nachhaltige und interoperable Verknüpfung von Forschungsdaten erwähnt, insbesondere bei der Modellierung und Annotation historischer Quellen. Matthias WEBER geht noch einen Schritt weiter und zeigt, wie graphbasierte Datenbanken relational-transitive Verbindungen, beispielsweise zwischen Personen, Orten und Ereignissen, modellieren können. Der Beitrag verdeutlicht, dass durch die Integration von Normdaten wie ‚Wikidata‘-IDs neue interdisziplinäre Forschungsperspektiven eröffnet und historische Netzwerke detaillierter analysiert werden können. Diese Ansätze veranschaulichen, wie Normdaten die Basis für nachhaltige und innovative Forschungspraktiken schaffen.

Schließlich wird auch die Entwicklung neuer Methoden für die Textanalyse in den Beiträgen des Themenhefts reflektiert. Digitale Verfahren zur Textanalyse

entwickeln und diskutieren TOMASEK, HART, SCHULTE und REUL, die sich mit Wiederholungsstrukturen im ‚Marienleben‘ Philipps von Seitz beschäftigen. Sie zeigen, wie für die Messbarmachung von Wiederholungen Besonderheiten der historischen Sprache berücksichtigt werden können (und müssen). Das entwickelte Verfahren zur Erkennung von Wiederholungen wird zunächst für die Analyse des ‚Marienlebens‘ eingesetzt und in einem zweiten Schritt auf ein größeres Korpus angewandt, um den Befund gattungsgeschichtlich einzuordnen.

Die Beiträge des Themenhefts repräsentieren nicht nur die disziplinäre Vielfalt der digitalen Mediävistik, sondern auch die Vielfalt der methodischen Zugriffe, die dieses Forschungsfeld prägen. Sie zeigen, wie traditionelle geisteswissenschaftliche Fragestellungen durch digitale Verfahren erweitert und vertieft werden können und wie innovative Ansätze neue Perspektiven auf historische Quellen und ihre Kontextualisierung eröffnen. Gleichzeitig stehen die Beiträge für eine transparente und offene Wissenschaft, die durch den Einsatz von Normdaten, interoperablen Standards und frei zugänglichen Infrastrukturen eine nachhaltige Nachnutzung ermöglicht. Die kollaborative und innovative Forschung, die hier sichtbar wird, verdeutlicht, dass die digitale Mediävistik nicht nur bestehende Methoden aufgreift und anpasst, sondern als Schnittstelle für neue interdisziplinäre Verknüpfungen agiert. In diesem Sinne verstehen wir die in den Beiträgen vorgestellten Ansätze als Momentaufnahmen einer im Wandel begriffenen Arbeitspraxis, die durch methodische Offenheit zum einen interoperabel und zum anderen anschlussfähig für künftige Entwicklungen bleibt. Fünf Jahre nach Erscheinen des ersten Themenhefts zur digitalen Mediävistik ist unseres nun eine erneute Einladung an die mediävistische Community, die Methoden und Praktiken weiterzudenken, um den rasanten technologischen Wandel und die durch ihn hervorgebrachten Verfahren für die Mittelalterforschung fruchtbar zu machen und die damit erzielten Ergebnisse kritisch zu reflektieren.

## 8 Danksagung

Ein Themenheft entsteht in einer gemeinschaftlichen Anstrengung, die von vielen Schultern getragen wird. Wir danken den Autor\*innen des Bandes für die intensive Zusammenarbeit über einen Zeitraum von eineinhalb Jahren hinweg. Erst der fruchtbare Austausch im ‚Netzwerk Offenes Mittelalter‘ und die vielen Impulse zu den gemeinsamen Themen haben die Idee zu diesem Themenheft überhaupt Gestalt annehmen lassen. Allen Mitgliedern und Assoziierten sei deswegen für die bereichernden Diskussionen und Anregungen gedankt. Die drei Hilfskräfte des Netzwerks, Leonie WEISS, Julia HÖPFNER und Tabea RUDERSDORFER, haben nicht nur das Netzwerk bei organisatorischen Aufgaben unterstützt, sondern in der Endphase auch die Herausgeberinnen bei der Redaktion des Themenhefts.

Wir sind dafür sehr dankbar! Die DFG förderte das Netzwerk mit Mitteln für die Arbeitstreffen. Der Mediävistikverband übernahm einige Kosten für den Heftworkshop. Diese finanzielle Unterstützung erleichterte den produktiven Austausch sehr und wir danken beiden Förderern dafür.

Besonders herzlich danken wir Hannah BUSCH, deren unschätzbare redaktionelle Mitarbeit sich auf alle wesentlichen Entwicklungsstufen des vorliegenden Themenheftes erstreckte, und Isabelle MANDRELLA, die die Entstehung des Heftes als Schriftführerin des Mediävistikverbandes sehr umsichtig und äußerst tatkräftig begleitete.

## Literaturverzeichnis

- Antenhofer, Christina, Christoph Kühberger u. Arno Strohmeyer (Hgg.):** Digital Humanities in den Geschichtswissenschaften. Wien 2023. <https://doi.org/10.36198/9783838561165> (Zugriff: 15.12.2024).
- Beek, Alan L. van u. Luise Borek:** Workshop „Schnittstelle Mediävistik“ im Rahmen des DFG-Netzwerks „Netzwerk Offenes Mittelalter“ zu Linked Open Data. In: *Mittelalter* (2023). <https://doi.org/10.58079/rhbp> (Zugriff: 15.12.2024).
- Bleier, Roman, Franz Fischer, Torsten Hiltmann u. a.:** Digitale Mediävistik und der deutschsprachige Raum. In: Dies. (Hgg.): *Digitale Mediävistik (Das Mittelalter. Perspektiven mediävistischer Forschung 24)*. Berlin 2019, S. 1–12. <https://doi.org/10.1515/mial-2019-0001> (Zugriff: 15.12.2024).
- Borek, Luise, Katharina Zeppezauer-Wachauer u. Nora Ketschik:** Eindeutig Uneindeutig. Zur Modellierung von Unschärfe in der Mediävistik. In: *Mittelalter* (2022). <https://doi.org/10.58079/rha5> (Zugriff: 15.12.2024).
- Burrichter, Brigitte, Björn Gebert, Christoph Mackert u. Gabriel Viehhauser:** Digitale Mediävistik. In: Wolfram Drews, Matthias Müller u. Regina Toepfer (Hgg.): *Mediävistik 2021. Positionen, Strategien, Visionen (Das Mittelalter. Perspektiven mediävistischer Forschung 26)*, Heidelberg 2021, S. 101–117. <https://doi.org/10.17885/heipub.mial.2021.1.24312> (Zugriff: 15.12.2024).
- Busch, Nathanael u. Diana Müller:** Normdaten in den Geisteswissenschaften. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 53 (2023), S. 781–796. <https://doi.org/10.1007/s41244-023-00295-1> (Zugriff: 15.12.2024).
- CLARIAH-AT Konsortium:** Digital Humanities Austria Strategie 2021+. Vier Leitlinien für Digital Humanities in Österreich. <http://gams.uni-graz.at/o:clariah.dha-strategie-2021> (Zugriff: 15.12.2024).
- Genicot, Léopold:** Pour une organisation de la recherche en histoire médiévale, II. L’information sur les recours aux ordinateurs. In: *Francia* 3 (1975), S. 643–647.
- Haaren, Suzette van u. Juliet Diekkämper:** List of Digital Resources for Medieval Studies in German-speaking Countries. Zenodo 2024. <https://doi.org/10.5281/zenodo.13332123> (Zugriff: 15.12.2024).
- Kestemont, Mike, Folgert Karsdorp, Elisabeth de Bruijn u. a.:** *Forgotten*

Books: The Application of Unseen Species Models to the Survival of Culture. In: *Science* 375/6582 (2022), S. 765–769. <https://doi.org/10.1126/science.abl7655> (Zugriff: 15.12.2024).

**Ketschik, Nora u. Max Grüntgens:**

Methodenkritik. Einblicke und Lessons Learned im Umgang mit Linked Open Data. In: *Mittelalter* (2024). <https://doi.org/10.58079/11pzx> (Zugriff: 15.12.2024).

**Klug, Helmut W. (Hg.)** unter Mitarbeit von Selina Galka u. Elisabeth Steiner im HRSM-Projekt „Kompetenznetzwerk Digitale Edition“: KONDE Weißbuch. <https://www.digitale-edition.at> (Zugriff: 15.12.2024).

**Lejeune, Edgar:** „Open“ or „Close“ Research Instruments? Conflicting Rationales in the Organization of Early Digital Medieval History in Europe (1960–1990). In: *Digital Humanities Quarterly* 18/2 (2024). <https://dhq-static.digital>

[humanities.org/pdf/000741.pdf](https://humanities.org/pdf/000741.pdf) (Zugriff: 15.12.2024).

**Rapp, Andrea:** Eine digitale Geschichte der germanistischen Mediävistik. Keynote des Workshops „Schnittstelle Mediävistik“ vom 3. und 4. April 2023. Salzburg. <https://unitv.org/beitrag.asp?ID=1142> (Zugriff: 15.12.2024).

**Zeppezauer-Wachauer, Katharina u.**

**Marco Heiles:** Eine digitale Textreihentypologie für deutschsprachige Texte des Mittelalters und der Frühen Neuzeit. In: *Mittelalter* (2023). <https://doi.org/10.58079/rhbs> (Zugriff: 15.12.2024).

**Zeppezauer-Wachauer, Katharina, Lina**

**M. Zangerl u. Johannes Stigler:** Nachhaltigkeit. Schlüsselkriterium für die Akzeptanz des digitalen Weges in den Geisteswissenschaften. In: Antenhofer, Kühberger u. Strohmeyer 2024, S. 579–594. <https://doi.org/10.36198/9783838561165> (Zugriff: 15.12.2024).

# Historisierte Ontologien für *Linguistic Linked Open Data*-Ressourcen des Mittelalters

**Abstract** Historical linguistic resources store valuable information about words and their meanings. These resources often suffer from limited online accessibility due to varied data formats, historical language stages, cultural contexts, etc. Modelling these resources as Linked Open Data (LOD) can overcome these challenges. Linguistics increasingly creates LOD using domain-specific ontologies for labelling linguistic elements such as words and multi-word expressions. For data access based not (only) on words but on their meanings, one must link these meanings to an extra-linguistic knowledge base describing the things of the world (lexical-semantic mapping, LSM). This enables a lexical-semantics-based data retrieval that is essential for asking cultural-historical questions of linguistic resources across various language stages. However, knowledge bases are typically ahistorical, reflecting the modern world; but historical concepts require historicised ontologies capturing the specificity of historical explanation patterns to avoid anachronistic mappings. Thus, the development of historicised ontologies, as demonstrated with ontologies for medieval medicine and law conducted by the research project ALMA, is crucial. Used for the LSM and LOD modelling of historical linguistic resources, these ensure the temporal and cultural context of the data is preserved, facilitating accurate and meaningful data queries.

**Keywords** Digital Medieval Studies; Historicised Ontology; Linked Data; Linguistic Resources; Lexical-Semantic Mapping

## Kontakt

PD Dr. Sabine Tittel,  
Heidelberger Akademie  
der Wissenschaften, Projekt ALMA,  
Seminarstraße 3, 69117 Heidelberg,  
sabine.tittel@hadw-bw.de  
 <https://orcid.org/0000-0003-4746-7604>

## 1 *Linked Open Data* und deren Bedeutung für historische linguistische Ressourcen

In historischen linguistischen Ressourcen<sup>1</sup> sind wertvolle Ergebnisse der historischen lexikalischen Semantik zu – grob vereinfachend – ‚Wörtern‘ und ihren ‚Bedeutungen‘ gespeichert. Zwar wurden im Lauf der letzten Jahrzehnte zahlreiche historische linguistische Ressourcen retrodigitalisiert oder genuin digital im *World Wide Web* (WWW) zur Verfügung gestellt,<sup>2</sup> diese stellen jedoch Datensilos dar: Ein Datensilo macht digitale Ressourcen zwar öffentlich zugänglich, aber dies nur in einem proprietären Format und davon abhängigen Recherchefunktionen, oder nur mit eingeschränkten Zugangsmöglichkeiten.<sup>3</sup> Der Zugriff auf solche Daten leidet entlang mehrerer Dimensionen: Datenformate und -qualität, Forschungsfragen und -methoden, zeitlich-kulturelle Kontexte, Sprachen und historische Sprachstufen etc. sind heterogen und daher über die typischen Funktionen des WWW nur im Ausnahmefall gemeinsam und semantisch sinnvoll zu recherchieren.<sup>4</sup>

Die Modellierung der Daten gemäß dem Paradigma von *Linked Open Data* (LOD) für die Integration ins *Semantic Web*<sup>5</sup> ermöglicht einen Datenzugriff, der die Schwierigkeiten des Zugangs zu den Datensilos des WWW überwinden kann. Das LOD-Paradigma umfasst Methoden, Datenformate und Ontologien,<sup>6</sup> womit die über das WWW verteilten, isolierten und heterogenen Daten eindeutig und maschinenlesbar semantisch beschrieben, geteilt und mit anderen Daten verknüpft werden.<sup>7</sup> Das Standardformat für die Darstellung von LOD ist die Ontologiesprache *Resource Description Framework* (RDF).<sup>8</sup> RDF repräsentiert Informationen wie <Das Mittelalter> <ist eine> <Zeitschrift> oder <artère> <ist ein> <Nomen> in Form von Tripeln, i. e. gerichteten Graphen mit einem ‚Subjekt‘, das über

---

1 Eine linguistische Ressource ist ein Datensatz mit Sprachdaten; dazu gehören einsprachige oder mehrsprachige digitale Wortlisten und Wörterbücher, Texteditionen und -korpora, Grammatiken etc. Vgl. Lezcano, Sánchez-Alonso u. Roa-Valverde 2013, S. 264.

2 Vgl. die Entwicklungen in der digitalen Geschichtswissenschaft. Einen Überblick dazu gibt es in Romein, Kemman, Birkholz u. a. 2020.

3 Vgl. McCrae, Montiel-Ponsoda u. Cimiano 2012, S. 25.

4 Mit Bezug zu historischen linguistischen Ressourcen näher ausgeführt in Tittel 2023, S. 154–158.

5 Das *Semantic Web* ist eine Erweiterung des WWW, das die Informationen des WWW mit Bedeutung versieht, maschinenlesbar und interoperabel macht; vgl. Berners-Lee, Hendler u. Lassila 2001.

6 Eine Ontologie in der Informatik ist ein formales Repräsentationssystem, das Entitäten/Begriffe (als Klassen) eines Wissens- oder Gegenstandsbereichs und die zwischen ihnen bestehenden Relationen maschinenlesbar beschreibt; vgl. Grim, Abecker, Völker u. a. 2011, S. 510.

7 Vgl. Bizer, Heath u. Berners-Lee 2009.

8 <https://www.w3.org/TR/rdf11-concepts/> (Zugriff: 10.03.2025).

ein ‚Prädikat‘ auf ein ‚Objekt‘ verweist.<sup>9</sup> Dabei werden Subjekte, Prädikate und Objekte in RDF durch URLs, die eine Maschine auffinden und verarbeiten kann, eindeutig beschrieben. Das Tripel <artère> <ist ein> <Nomen> wird damit zu <<https://cnrtl.fr/definition/artère>> <<http://www.lexinfo.net/ontology/2.0/lexinfo#partOfSpeech>> <<http://www.lexinfo.net/ontology/2.0/lexinfo#Noun>>, wobei das Subjekt auf die URL des Eintrags zu französisch *artère* im Wörterbuch ‚Trésor de la langue française informatisé‘ (TLFi)<sup>10</sup> verweist und Prädikat und Objekt von der Domänenontologie für linguistische Termini ‚lexinfo‘<sup>11</sup> bereitgestellt werden. Objekte können auch Zeichenketten, sogenannte Literale, sein, die mit einem ISO-639-Sprachcode<sup>12</sup> versehen werden, wie zum Beispiel in <<http://www.lexinfo.net/ontology/2.0/lexinfo#Noun>> <rdfs:label> <“Substantiv“@de>, wo über das Prädikat label, bereitgestellt von RDFS,<sup>13</sup> die Verbindung des Subjekts ‚Noun‘ zum deutschsprachigen Literal ‚Substantiv‘ geschaffen wird.

Die Vorteile von LOD umfassen den einheitlichen Datenzugriff über HTTP, die Ressourcenintegration über die Datenverknüpfungen, die strukturelle Interoperabilität durch RDF und die konzeptionelle Interoperabilität durch den gemeinsamen Einsatz von Ontologien.<sup>14</sup>

Im WWW sind inhaltliche Zusammenhänge zwischen Informationen unterschiedlicher Ressourcen zwar implizit vorhanden, aber für den menschlichen Benutzer in der immensen Informationsflut nur schwer erkennbar. Als LOD werden diese Zusammenhänge explizit gemacht. Dies – zusammen mit der Datenanfrage über die Disziplinen, Inhalte, Formate und Sprachen der heterogenen Ressourcen hinweg – eröffnet neue Möglichkeiten der Wissensgenerierung. In Bezug auf die digitalen Geschichtswissenschaften schreiben ROMEIN, KEMMAN, BIRKHOLZ u. a.: „For historical scholarship, this means that statements about a single entity can be taken from a large amount of sources spread over many archives in order to gain a bigger picture or to identify opposing views.“<sup>15</sup> LOD entsprechen den FAIR-Prinzipien *Findable, Accessible, Interoperable, Reusable*.<sup>16</sup>

---

<sup>9</sup> Die Bezeichnungen gehören zur Syntax von Programmiersprachen und lehnen sich an die der Syntax der Linguistik an.

<sup>10</sup> <http://atilf.atilf.fr/>, <https://cnrtl.fr/definition/artère> (Zugriff: 10.03.2025).

<sup>11</sup> <http://lexinfo.net/> (Zugriff: 10.03.2025).

<sup>12</sup> [https://iso639-3.sil.org/code\\_tables/639/data](https://iso639-3.sil.org/code_tables/639/data) (Zugriff: 10.03.2025).

<sup>13</sup> *Resource Description Framework Schema*, eine weitere Ontologiesprache, ausdrucksstärker als RDF; <https://www.w3.org/TR/rdf-schema/> (Zugriff: 10.03.2025).

<sup>14</sup> Vgl. Chiarcos, McCrae u. Cimiano 2013; Tittel 2024, S. 105–107.

<sup>15</sup> Romein, Kemman, Birkholz u. a. 2020, S. 301 f.

<sup>16</sup> <https://www.go-fair.org/fair-principles/> (Zugriff: 10.03.2025).

Zunehmend werden von der historischen Linguistik *Linguistic Linked Open Data* (LLOD) erstellt.<sup>17</sup> Für die konzeptuelle Interoperabilität der linguistischen Daten sind bereits grundlegende Ontologien in der Disziplin etabliert. Diese stellen Klassen und Relationen bereit, die etwa die Auszeichnung von Zeichenketten als ein ‚Wort‘ (zum Beispiel ‚Schach‘), als ‚Mehrwortverbindung‘ (‚jemanden schachmatt setzen‘), als ‚Wortklasse‘ (zum Beispiel ‚Verb‘) etc. ermöglichen. Die zentrale Ontologie für LLOD ist ‚OntoLex-Lemon‘.<sup>18</sup>

## 2 Lexikalisch-semantisches Mapping

Wörter und ihre Bedeutungen sind Bestandteile des einzelsprachlichen Zeichens im semiotischen Fünfeck der Linguistik, das für unsere Herangehensweise die wissenschaftstheoretische Basis bildet: Das einzelsprachliche Zeichen besteht aus einer Ausdrucksseite, das heißt der (lautlichen oder, für historische Daten, schriftlichen) sprachlichen Repräsentation, und aus einer Inhaltsseite, das heißt der Bedeutung. Letztere verweist auf ein außersprachliches, abstraktes Konzept und dieses wiederum auf einen konkreten Referenten in der Wirklichkeit.<sup>19</sup> Altfranzösisch *artere* beispielsweise verweist auf das mittelalterliche Konzept der ‚Arterie‘, das Konzept schließlich auf die Blutbahn.

Während mit ‚OntoLex-Lemon‘ Informationen zur Ausdrucksseite des sprachlichen Zeichens – die historischen Schreib- und Zitierformen der Wörter, ihre dialektalen Varianten, die Bestandteile einer Mehrwortverbindung etc. – sehr gut als LOD modelliert werden können, besteht der zentrale, aber zeitintensive Aspekt des Modellierens darin, einen Zugang zu den Daten über die Inhaltsseite des sprachlichen Zeichens zu schaffen, um darüber auf die außersprachlichen Konzepte und die dahinter liegenden konkreten Dinge der Welt zugreifen, sie verknüpfen und damit kontextualisieren zu können. Denn es ist genau dieser bedeutungsbasierte Datenzugriff, der für kulturell-historische Fragestellungen an linguistische Ressourcen essentiell ist: Er ermöglicht Antworten, die nicht (oder nicht nur) auf der Ausdrucksseite des sprachlichen Zeichens, sondern auf dessen Inhaltsseite basieren, und dies konzeptbasiert und damit außerhalb des einzelsprachlichen Zeichens, also über verschiedene Sprachen und ihre historischen Sprachstufen hinweg.

Wenn der Nutzer beispielsweise nach Dingen im Zusammenhang mit dem Thema ‚Blutfluss‘ in verschiedenen Zeiten und Kulturen recherchiert, verlangt

---

17 Für Beispiele vgl. Tittel 2023, S. 163–165.

18 <https://www.w3.org/2016/05/ontolex/> (Zugriff: 10.03.2025). Einen Überblick über Technologien, Ontologien und Methoden geben Cimiano, Chiarcos, McCrae u. a. 2020; Khan, Chiarcos, Declerck u. a. 2022.

19 Vgl. Blank 2001, S. 9.

ein bedeutungsbasierter Datenzugang dem Nutzer für die Anfrage an die Daten eben nicht ab zu wissen, dass die altfranzösische sprachliche Realisierung – das sprachliche Zeichen – für das Konzept der ‚Arterie‘ das Wort *artere* ist. Dies mag im Falle von *artere* zwar fast trivial erscheinen, wenn man aber verschiedene altfranzösische Bezeichnungen für Blutbahnen betrachtet, wird die Problematik eines auf der Ausdrucksseite des sprachlichen Zeichens beruhenden Zugriffs deutlich: altfranzösisch *rame* m./f. („Subdivision einer Blutbahn“), *raincel* m. (id.), *guidege* subst. („seitlich am Hals gelegene Blutbahn, die das Blut zum Kopf transportiert“), *veine kilis* f. („Blutbahn, die das ‚ernährende‘ Blut von der Leber in die Körperteile transportiert“), *dilatoire* adj. („fähig sich auszudehnen [von einem Blutgefäß des Skrotums]“) etc.<sup>20</sup>

Die Voraussetzung für einen solchen Datenzugriff ist das lexikalisch-semantische Mapping (LSM).<sup>21</sup> Dieses verknüpft die Inhaltsseite des sprachlichen Zeichens in LLOD-Ressourcen mit einem Begriff einer Ontologie, i. e. einer externen und sprachunabhängigen Wissensbasis, die die Dinge der Welt abbildet: die Arterie, das Blutgefäß, Gesundheit, Krankheit. ‚OntoLex-Lemon‘ bietet die Klassen zur Modellierung der Inhaltsseite des sprachlichen Zeichens mit ‚Bedeutungsdefinition‘ (LexicalSense) und dem zugrunde liegenden außersprachlichen Konzept (LexicalConcept)<sup>22</sup> (übereinstimmend mit dem semiotischen Fünfeck) und die nötigen Relationen (reference; isConceptOf), um diese Klassen mit einer Entität einer externen Wissensbasis zu verknüpfen.

Aber worauf mappen? Für zeitgenössische Lebenswelten stehen mit ‚Dbpedia‘, einer multilingualen Enzyklopädie für das *Semantic Web* und zentralem Knotenpunkt in der LOD-Landschaft, und ‚Wikidata‘, einer Wissensdatenbank, die unter anderem ‚Wikipedia‘ und ‚Dbpedia‘ strukturierte Daten zur Verfügung stellt,<sup>23</sup> solche Wissensbasen zur Verfügung.<sup>24</sup> Allerdings sind beide, wenn auch enzyklopädisch und sprachunabhängig, für das LSM historischer Konzepte ungenügend: Sie sind universalistisch und ahistorisch ausgerichtet und bilden in der Regel die Erkenntnisse moderner Wissenschaftsdisziplinen ab.

Zum Beispiel ist das Konzept der ‚Arterie‘ das der modernen Anatomie, welches die Blutbahn als Teil der Blutzirkulation versteht. Aber bevor 1628 William Harveys Beschreibung des Blutkreislaufs<sup>25</sup> das Verständnis der Funktionen aller

20 Vgl. Tittel 2004, S. 231–233, 244–246, 252, 256.

21 Vgl. ebd., Kap. 18.

22 <https://www.w3.org/2016/05/ontolex/> (Zugriff: 10.03.2025).

23 <https://www.wikidata.org/>; <https://www.dbpedia.org/> (Zugriff: 10.03.2025).

24 Für eine ausführliche Evaluierung der Qualität (Glaubwürdigkeit, Reputation, Objektivität und Überprüfbarkeit) von ‚Dbpedia‘ und ‚Wikidata‘ als Wissensbasen für das Semantic Web vgl. Färber, Bartscherer, Menne u. a. 2017, S. 19–46.

25 Vgl. Schipperges 1990.

mit dem Blutfluss verknüpften Dinge grundlegend korrigierte, war dieses weitestgehend von den antiken Vorstellungen der Pneumalehre<sup>26</sup> geprägt. Ein Mapping von etwa altfranzösisch *artere* auf den ‚Dbpedia‘-Eintrag ‚Artery‘ wäre damit ahistorisch. Generell wäre ein LSM historischer Konzepte ‚alter Wörter‘ auf ‚moderne Konzepte‘ durchzogen von semantischen Brüchen, die eine sinnvolle Anfrage der Daten aus verschiedenen zeitlichen und kulturellen Kontexten in vielen Fällen unmöglich machen würde. Ein historisch korrektes Mapping ist mangels historisch definierter Einträge in einer Wissensbasis sehr oft nicht möglich; ein Mapping von Konzepten, die gar nicht mehr Bestandteil des modernen Lebens sind, ebenfalls nicht.

### 3 Historisierte Ontologien

Es ist daher nötig, historisierte Ontologien einzusetzen, die der Spezifik mittelalterlicher Erklärungsmuster gerecht werden und anachronistische Zuordnungen verhindern. Die Ontologien müssen die erkenntnistheoretischen Kontinuitätsbrüche zwischen mittelalterlichen und aktuellen Wissens- und Erklärungssystemen, zu denen es – etwa in Humanismus und Aufklärung – kam, identifizieren und über ihre Klassen und Relationen nicht moderne, sondern mittelalterliche Wissenssysteme abbilden. Durch den sprachunabhängigen und historisierenden Ansatz können Ontologien einen signifikanten Beitrag zur Forschungslandschaft von LOD leisten und der Nutzerkreis der Ontologien kann weit über den speziellen Bereich der historischen Sprachforschung hinaus auf alle historisch arbeitenden Wissenschaften erweitert werden.<sup>27</sup>

#### 3.1 Historische Ontologien der Philosophie

Ebenso wie die Ontologie im Sinne der Informatik ist auch die historische bzw. historisierte Ontologie in Bezug zu setzen zur Ontologie der Philosophie. Das Forschungsfeld der historischen Ontologie wird innerhalb der Disziplin der Philosophie unterschiedlich definiert.<sup>28</sup> Unsere Herangehensweise, die in der Kombination mit der historischen lexikalischen Semantik gründet, wird philosophietheoretisch untermauert durch die Definition Ian HACKINGS, der die historische Ontologie als eine historisierte und sprachanalytisch gestützte Ausprägung einer

---

26 Vgl. Putscher 1974.

27 Zu Ontologien für mittelalterliche Manuskriptbeschreibungen siehe auch Bauer, Bleier u. Sonnberger in diesem Heft (insb. Kap. 3.2).

28 Für einen Überblick vgl. Lindén 2019, S. 12–23.

philosophischen Begriffsanalyse versteht.<sup>29</sup> Der Terminus ‚historische Ontologie‘ wird als Sammelbegriff für Analysemethoden verwendet, die sich mit den Wechselbeziehungen zwischen historisch kontextualisierten Wörtern, Sätzen, Redeweisen und -stilen auf der einen Seite und Begriffen auf der anderen Seite befassen:

[A] concept is nothing other than a word in its sites. That means attending to a variety of types of sites: the sentences in which the word is actually [...] used, those who speak those sentences, with what authority, in what institutional setting [...]. The logical relations among them [the concepts] were formed in time, and they cannot be perceived correctly unless their temporal dimensions are kept in view.<sup>30</sup>

Die Tatsache, dass das Erfahren der Welt subjektiv ist und dass diese Tatsache epistemologisch zu berücksichtigen ist, ist bekannt. WEYL schreibt beispielsweise:

The objective world [...] with which we reckon continually in our daily lives and which the natural sciences attempt to crystallize by methods representing the consistent development of those criteria by which we experience reality in our natural everyday attitude—this *objective* world is of necessity *relative* [...]. It seems to me that this pair of opposites, subjective-absolute and objective-relative, contains one of the most fundamental epistemological insights which can be gleaned from science. Whoever desires the absolute must take the subjectivity and egocentricity into the bargain; whoever feels drawn toward the objective faces the problem of relativity.<sup>31</sup>

### 3.2 Historische und historisierte Ontologie im Sinne der Informatik

In Analogie zur historischen Ontologie der Philosophie stehen die historischen bzw. historisierten Ontologien der Informatik im Kontext des *Semantic Web*. MEROÑO-PEÑUELA, ASHKPOUR, ERP u. a. betonen, wie wichtig es für den Historiker sei, in der digitalen Arbeit semantisch aufbereitete Quellen nutzen zu können, die die historischen Daten sowohl zeitlich als auch regional (und ich füge hinzu: kulturell<sup>32</sup>) kontextualisieren; nötig dafür seien Datenmodelle – i. e. Ontologien

<sup>29</sup> Vgl. Hacking 2002, S. 24f.; Dittrich 2006, S. 36.

<sup>30</sup> Hacking 2002, S. 17–25; vgl. Dittrich 2006, S. 9.

<sup>31</sup> Weyl 1949, S. 116.

<sup>32</sup> Vgl. Lindén 2019, S. 22.

für das *Semantic Web* –, die den Daten Struktur verleihen und sie für eine effiziente Mensch-Maschine-Kommunikation vorbereiten.<sup>33</sup>

Begriffe der Welt in Ontologien für das *Semantic Web* zu historisieren, die Definition der außersprachlichen Konzepte also historisch zu differenzieren, kann prinzipiell nach zwei Methoden erfolgen: (i) Einerseits lassen sich historisierte Ontologien erarbeiten. Diese sind typischerweise als statisch konzipiert zu verstehen, das heißt, sie stellen eine Momentaufnahme eines gegebenen Zeitpunkts oder auch -abschnitts einer Domäne dar, und sie arbeiten mit Begriffsdefinitionen, die als die zu der gegebenen Zeit gültigen erkannt werden. Interoperabilität zwischen mehreren historisierten Ontologien lässt sich etablieren etwa durch das Verknüpfen mit einer *Upper Ontology*, zum Beispiel mit der Ontologie SUMO;<sup>34</sup> diese Ontologie dient als Brücke zu allgemeineren Strukturen und als zentrale Verankerung einzelner Ontologien. Auf unseren Ansatz, eine eigene Relation als ‚Brücke‘ explizit für die Verknüpfung mit einer modernen Entsprechung einzuführen, komme ich weiter unten zurück.

(ii) Andererseits lassen sich historische Ontologien erarbeiten. Diese sind als dynamisch zu verstehen, das heißt, sie bilden Begriffe und Relationen der Ontologie so ab, dass sie mehrere verschiedene, sich durch die Zeit hindurch verändernde Definitionen besitzen, siehe die theoretischen Ansätze in WANG, SCHLOBACH und KLEIN.<sup>35</sup> IDE und WOOLNER schlagen für eine historische Ontologie eine Architektur vor, die zwei Perspektiven bereitstellt: eine synchrone Sicht – dies entspricht der Momentaufnahme einer historisierten Ontologie – und eine diachrone Sicht, die aus der Verknüpfung der Momentaufnahmen entsteht.<sup>36</sup>

### 3.3 Für das lexikalisch-semantische Mapping von LLOD zur Verfügung stehende historische oder historisierte Ontologien

Viele Unternehmungen der Geschichtswissenschaften, die das Vernetzen historischer Information mit Hilfe von historisierten oder historischen Ontologien zum Ziel haben, fokussieren auf Orts- und Personennamen, auf Berufsbezeichnungen als demographische und sozial-ökonomische Indikatoren<sup>37</sup> und auf Ereignisse.<sup>38</sup> Für die Architektur, für kulturelle und andere Objekte des materiellen Kulturerbes

33 Vgl. Meroño-Peñuela, Ashkpour, Erp u. a. 2014.

34 ‚Suggested Upper Merged Ontology‘; vgl. Pease u. Niles 2001.

35 Vgl. Wang, Schlobach u. Klein 2011.

36 Vgl. Ide u. Woolner 2007, speziell S. 9f.

37 Siehe z.B. das Projekt ‚HISCO – History of Work‘; <https://iisg.amsterdam/en/data/data-websites/history-of-work> (Zugriff: 10.03.2025).

38 Vgl. Meroño-Peñuela, Ashkpour, Erp u. a. 2014, S. 11f., 14f., 17.

stehen zum Beispiel die Ontologien des ‚Getty Research Institute‘ als wertvolle Ressourcen bereit.<sup>39</sup> Eine maßgebliche Ontologie für die Abbildung von Daten des Kulturerbes, fokussiert auf das ‚Ereignis‘ (mit Akteur, Aktion, Zeit, Ort, Resultat etc.), ist das Datenmodell CIDOC CRM.<sup>40</sup>

Historische Personen, Orte, kirchliche Institutionen, Ämter, Wappen und Ähnliches werden als Wissensbasis in RDF vom Forschungsvorhaben ‚Wissens-Aggregator Mittelalter und Frühe Neuzeit‘ (WIAG)<sup>41</sup> erfasst. Für mittelalterliche Glasmalerei (in Deutschland bis 1550) erstellt das ‚Corpus Vitrearum Medii Aevi‘ (CVMA)<sup>42</sup> eine Ontologie zu form-, stil- und bildgeschichtlichen Aspekten der Kunstgeschichte. Die ‚Mittelhochdeutsche Begriffsdatenbank‘ (MHDBDB)<sup>43</sup> erarbeitet eine ‚Ontology of Narratives of the Middle Ages‘<sup>44</sup> für die Beschreibung von Handlungen, Akteuren und Schauplätzen im Text- und Bildmedium. Das zentrale Gerüst für die LOD-Modellierung und die Datenvisualisierung der MHDBDB ist ein allgemeines, als Baumstruktur organisiertes Begriffssystem für das Mittelhochdeutsche in RDF.<sup>45</sup> Dieses entstand ausgehend vom ‚Begriffssystem als Grundlage für die Lexikographie‘ von HALLIG und VON WARTBURG (Hallig-Wartburg),<sup>46</sup> das in das Datenmodell SKOS,<sup>47</sup> ein Ontologietypus mit geringer Komplexität, überführt und um zahlreiche Kategorien erweitert wurde. Hallig-Wartburg wurde 1952 als onomasiologisches, das heißt konzeptbasiertes Gerüst zur Kategorisierung der Dinge der Welt für die historische Sprachwissenschaft kreiert. Die außersprachlich gedachten Begriffe des Hallig-Wartburg bilden die Welt der ersten Hälfte des 20. Jahrhunderts ab.<sup>48</sup> Der Schritt von einem SKOS zum Ontologieformat OWL ist aufgrund von Inkonsistenzen in der Kategorisierungsstruktur des Hallig-Wartburg

---

39 ‚Art & Architecture Thesaurus‘ (AAT), ‚Cultural Objects Name Authority‘ (CONA); <https://www.getty.edu/research/tools/vocabularies/index.html> (Zugriff: 10.03.2025).

40 ‚Conceptual Reference Model‘; <https://www.cidoc-crm.org/> (Zugriff: 10.03.2025).

41 <https://adw-goe.de/forschung/weitere-forschungsprojekte/wissens-aggregator-mittelalter-und-fruehe-neuzeit/> (Zugriff: 10.03.2025).

42 <https://corpusvitrearum.de/> (Zugriff: 10.03.2025).

43 <https://mhdbdb.sbg.ac.at/> (Zugriff: 10.03.2025).

44 Siehe auf dem GitHub der MHDBDB: <https://github.com/Middle-High-German-Conceptual-Database/onama> (Zugriff: 10.03.2025).

45 <https://mhdbdb.sbg.ac.at/#/concept/tree> (Zugriff: 10.03.2025).

46 Vgl. Hallig u. von Wartburg 1963.

47 *Simple Knowledge Organization System*, <https://www.w3.org/TR/2009/REC-skos-reference-20090818/> (Zugriff: 10.03.2025).

48 Zu den Mängeln des Hallig-Wartburg in Bezug auf Vollständigkeit, Abbildung von Fachlichkeit und Historizität vgl. Tittel, Gillis-Webber u. Nannini 2020: Die Arbeit modelliert die Begriffe des Hallig-Wartburg in RDF und SKOS und zeigt mit ersten Schritten zur Überführung in die Web Ontology Language (OWL), wie Hallig-Wartburg in eine Ontologie im Sinne des Semantic Web weiterzuentwickeln ist. OWL ist eine Ontologiesprache, die RDF mit mehr Expressivität erweitert, siehe <https://www.w3.org/TR/2012/REC-owl2-primer-20121211/>

nicht einfach, aber wichtig, um mehr Expressivität für die Beschreibung der Relationen zwischen den Kategorien zu erreichen und darüber der geringeren konzeptuellen Abstraktion und den fehlenden Nuancen des semantischen Ausdrucks von SKOS entgegenzuwirken.

Semantische Netze gehören als Teilmenge der Thesauri zwar in das breite Spektrum von Ontologien,<sup>49</sup> sind aber nicht außersprachlich, sondern einzelsprachlich gebunden, indem sie die sprachlichen Zeichen einer natürlichen Sprache und ihre synonymischen, antonymischen, hyponymischen etc. Relationen in Netzstrukturen modellieren. Ein sehr bekanntes Beispiel, vielleicht der Goldstandard der semantischen Netze, ist WordNet,<sup>50</sup> eine Datenbasis des Englischen mit Substantiven, Verben, Adjektiven und Adverbien.

Der ‚Historical Thesaurus of English‘ (HTE)<sup>51</sup> ist ebenfalls ein einzelsprachlich basiertes System. Er klassifiziert den Wortschatz der englischen Sprache vom Altenglischen bis heute, wie er vom ‚Oxford English Dictionary‘<sup>52</sup> verzeichnet wird, und integriert ebenfalls das Vokabular des ‚Thesaurus of Old English‘.<sup>53</sup> Die über 40 Jahre währende Arbeit resultierte in rund 235.000 lexikalisch-semantische Kategorien, konstruiert aus den englischen Wörtern und ihren Bedeutungen. Trotz der Sprachgebundenheit ist HTE ein umfassendes Referenzwerk für historische Inhalte und von grundlegender Bedeutung für ein LSM.

### 3.4 Arbeiten zum lexikalisch-semantischen Mapping historischer linguistischer *Linked Open Data*

Das Problem des LSM historischer LLOD wird nach unserer Kenntnis in der Literatur nur wenig diskutiert. BOZZI beschreibt die Unmöglichkeit, WordNet für ein LSM der altokzitanischen medizinisch-botanischen Termini des Projekts ‚Dictionnaire de Termes Médico-botaniques de l’Ancien Occitan‘ (DiTMAO) einzusetzen.<sup>54</sup> DECLERCK, WANDL-VOGT und MÖRTH (zum ‚Wörterbuch der bairischen Mundarten in Österreich‘) und BELLANDI, GIOVANNETTI, PICCINI u. a. 2017 (zu DiTMAO sowie zu ‚Un viaggio virtuale attraverso l’Atlante di Matteo Ricci

---

(Zugriff: 10.03.2025). Zu den konzeptuellen Unterschieden zum SKOS der MHDBDB vgl. ebd., S. 6.

49 Vgl. z.B. Guarino, Oberle u. Staab 2009, S. 13.

50 <https://wordnet.princeton.edu/> (Zugriff: 10.03.2025).

51 Vgl. Kay, Roberts, Samuels u. a. 2009.

52 <https://www.oed.com/> (Zugriff: 10.03.2025).

53 Vgl. Roberts, Kay u. Grundy 2000.

54 Vgl. Bozzi 2016.

[1602]‘) deuten die Problematik nur an.<sup>55</sup> Sichtbar werden die Schwierigkeiten in den Bemühungen beim Projekt ‚Verba Alpina‘<sup>56</sup>, Konzepte der Kulturgeschichte des Alpenraums auf Einträge von ‚Wikidata‘ als außersprachliches Referenzsystem zu mappen.<sup>57</sup> TITTEL und CHIARCOS identifizieren im Zusammenhang mit der Modellierung des altfranzösischen Wörterbuchs ‚Dictionnaire étymologique de l’ancien français‘ (DEAF)<sup>58</sup> als LOD zwar das Problem als „historical semantic gap“ zwischen ‚alten‘ und ‚modernen‘ Konzepten, geben aber keine Lösung.<sup>59</sup> NANNINI präsentiert Arbeiten zum LSM der Daten des italienischen Wörterbuchs ‚Lessico Etimologico Italiano‘ (LEI).<sup>60</sup> Er prüft sowohl Hallig-Wartburg als auch HTE als Zielressourcen des Mappings und kommt zu dem Schluss, (i) dass das Mapping auf die feingranularen Einträge des HTE in vielen Fällen semantisch sinnvoller ist als auf die weniger tief hierarchisch strukturierten des Hallig-Wartburg und (ii), dass ein aus einer natürlichen Sprache (hier: Englisch) abgeleiteter Thesaurus auch zur Klassifizierung anderer Sprachvarietäten verwendet werden kann.<sup>61</sup> Die Problematik der Historizität wird aber nur angedeutet über den Verweis auf den *historical semantic gap* und auf die Diskussion der Mängel des Hallig-Wartburg in TITTEL, GILLIS-WEBBER und NANNINI.<sup>62</sup>

#### 4 Auf dem Weg zu Domänenontologien der mittelalterlichen Medizin und des Rechts

Historisierte Ontologien zu den beiden Domänen ‚Medizin‘ und ‚Recht‘ zu erarbeiten, hat sich das Projekt ‚Wissensnetze in der mittelalterlichen Romania‘ (ALMA)<sup>63</sup> zur Aufgabe gemacht. Im Rahmen der linguistischen Langzeitforschung werden zwei domänenspezifische Textkorpora mit Texten der mittelalterlichen Gallo- und Italomania (Französisch, Okzitanisch, Gaskognisch, Italienisch; 12.–15. Jahrhundert) erstellt. Auf dieser empirischen Grundlage werden die Domänenontologien entwickelt. Diese bilden als Klassen die Konzepte ab, die sich durch die Analyse

---

55 Vgl. Declerck, Wandl-Vogt u. Mörth 2015, S. 348–353; <https://wboe.oeaw.ac.at/> (Zugriff: 10.03.2025); Bellandi, Giovannetti, Piccini u. a. 2017.

56 <https://dx.doi.org/10.5282/verba-alpina> (Zugriff: 10.03.2025).

57 Vgl. Tittel 2024, S. 145 f.

58 Vgl. Baldinger, Möhren u. Städtler 1971–2000.

59 Vgl. Tittel u. Chiarcos 2018.

60 Vgl. Pfister, Prifti u. Schweickard 1979–.

61 Vgl. Nannini 2023, S. 138.

62 Vgl. Tittel, Gillis-Webber u. Nannini 2020.

63 <https://www.hadw-bw.de/alma> (Zugriff: 10.03.2025).

der Fachtexte<sup>64</sup> und während der Ausarbeitung von lexikalisch-semantischen Studien – quasimonographische, sprachübergreifende, konzeptorientierte Wortschatzuntersuchungen – zu den Wörtern und ihren Bedeutungen identifizieren lassen.<sup>65</sup>

Im Rahmen des Projekts ALMA ist es nicht das Ziel, historische Ontologien wie beschrieben von IDE und WOOLNER (siehe oben) zu entwickeln, das heißt, die Begriffe vom Mittelalter bis heute mit all ihren konzeptuellen Veränderungen zu erfassen und zeitlich festzulegen. Anschlussfähig an eine potentielle historische Ontologie sind die Domänenontologien des Projekts aber in dem Sinne, als sie als Ressourcen nachgenutzt und in ein System von „trans-ontological relations“<sup>66</sup> integriert werden können. Sie können somit solide erarbeitete und empirisch gestützte Momentaufnahmen – als historische Ontologien im Sinne der Philosophie – bereitstellen, mit einem Zeitintervall von immerhin bis zu vier Jahrhunderten, je nach Konzept. Die Klassen und Relationen der Ontologien werden bei der *Linked Data*-Modellierung der Forschungsergebnisse des Projekts (der Korpus Texte, der lexikalisch-semantischen Studien und einer Forschungsbibliographie) eingesetzt, um die Strukturen der Wissensnetze erkennbar und abfragbar zu machen.

Es folgt ein Codebeispiel aus der Domäne der Medizin, das das Modellieren der Relationen `withinContext`, `hasModernCounterpart` (Abb. 1) und die Klasse `Artery` (Abb. 2) zeigt.<sup>67</sup> Die Metasprache der Ontologien von ALMA ist Englisch, die Definition der Klassen und Relationen erfolgt viersprachig (Englisch, Deutsch, Französisch, Italienisch) mit Hilfe von `rdfs:comment` (Abb. 1, Z. 17 f., 24–26); die Namen der Relationen werden über `rdfs:label` lesbar gemacht („to provide a human-readable version of a resource’s name“<sup>68</sup>) (Abb. 1, Z. 19, 27; Abb. 2, Z. 8).<sup>69</sup>

---

64 Die Textanalyse erfolgt sowohl mit den quantitativ-statistischen Methoden der Korpuslinguistik als auch mit den qualitativ-heuristischen Methoden der historischen Lexikologie. Die Ergebnisse werden abgesichert durch den Einbezug des (Mittel-)Lateinischen sowie der Hispanoromania und flankiert von den Geschichtswissenschaften.

65 Für eine Beispielstudie zum Konzept ‚Bauchfell‘ vgl. Prifti, Selig, Schweickard u. a. 2023, S. 319–327. *Nota bene*: Diese ist in Bezug auf die zukünftige optische Präsentation der Inhalte nicht verbindlich.

66 Ide u. Woolner 2007, S. 13.

67 *Nota bene*: Im Beispiel nicht berücksichtigt ist die Verknüpfung mit einer *Upper Ontology*. Da die Konzipierung der Ontologie noch nicht abgeschlossen ist, besitzen alle Angaben noch eine gewisse Vorläufigkeit: der Ontologienname ‚Ontology for the Middle Ages‘, das Namenspräfix ‚medieval‘ etc. Die Ontologie wird auf <https://lod.academy> oder <https://alma.hadw-bw.de> (Zugriff: 10.03.2025) zugänglich gemacht werden; ich verwende im Code aber den Namensraum `<https://example.org>`.

68 [https://www.w3.org/TR/rdf-schema/#ch\\_label](https://www.w3.org/TR/rdf-schema/#ch_label) (Zugriff: 10.03.2025).

69 Aus Platzgründen beschränke ich mich auf das Englische, sowohl für `comment` als auch für `label`.

```

1 <rdf:RDF xmlns:medieval="https://example.org/medieval#"
2   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
4   xmlns:owl="http://www.w3.org/2002/07/owl#"
5   xmlns:dct="http://purl.org/dc/terms/"
6   xmlns:vann="http://purl.org/vocab/vann/"
7   xmlns:time="http://www.w3.org/2006/time#">
8
9 <owl:Ontology rdf:about="https://example.org/medieval#">
10 <dct:description xml:lang="en">An ontology for the
11 description of knowledge domains of the Middle Ages.
12 </dct:description>
13 <dct:title xml:lang="en">Ontology for the Middle Ages
14 </dct:title>
15 <vann:preferredNamespacePrefix>medieval
16 </vann:preferredNamespacePrefix>
17 </owl:Ontology>
18
19 <!-- object properties -->
20 <!-- https://example.org/medieval#withinContext -->
21 <owl:ObjectProperty rdf:about="https://example.org/
22   medievalMedicine#withinContext">
23 <owl:inverseOf rdf:resource="https://example.org/
24   medievalMedicine#explainsByTheorie"/>
25 <rdfs:comment xml:lang="en">The 'withinContext'
26 property indicates a particular theoretical,
27 ideological, explanatory approach that is pertinent
28 to understand a notion.</rdfs:comment>
29 <rdfs:label xml:lang="en">within theorie</rdfs:label>
30 </owl:ObjectProperty>
31
32 <!-- https://example.org/medieval#hasModernCounterpart -->
33 <owl:ObjectProperty rdf:about="https://example.org/
34   medieval#hasModernCounterpart">
35 <rdfs:comment xml:lang="en">The 'modernCounterpart' property
36 indicates the correspondence to a modern term whose notion
37 is related, e.g., as a further development, advancement, or
38 derivative, but not congruent with the medieval notion of
39 the term.</rdfs:comment>
40 <rdfs:label xml:lang="en">has modern counterpart</rdfs:label>
41 </owl:ObjectProperty>
42 ...
43 </rdf:RDF>

```

**Abb. 1** | Auszug aus der Ontologie mit zwei Relationen withinContext und hasModernCounterpart. CC0.

Ziel ist eine stark axiomatisierte Ontologie,<sup>70</sup> über deren Relationen nicht nur hierarchische Klasse-Unterklasse-Beziehungen (Abb. 2, Z. 9) abgebildet werden sollen, sondern auch assoziative und für die Domäne spezifische Beziehungen, um nuanciert Klassen miteinander zu verknüpfen (Abb. 1, Z. 14, 22). Mit der

<sup>70</sup> Vgl. Grimm, Abecker, Völker u. a. 2011, S. 522–525.

```

1 <!-- classes -->
2 <!-- https://example.org/medieval#Artery -->
3 <owl:Class rdf:about="https://example.org/medieval#Artery">
4 <rdfs:comment xml:lang="en">Bloodvessel starting from the
5 heart in which the 'spiritualizing blood', enriched with
6 the 'spiritus vitalis' ('pneuma zotikon'), is distributed
7 to all parts of the body and flows back to the heart, thus
8 regulating the temperature of the body and sustaining
9 life.</rdfs:comment>
10 <rdfs:label xml:lang="en">artery</rdfs:label>
11 <rdfs:subClassOf
12   rdf:resource="https://example.org/medieval#BloodVessel"/>
13 <medieval:withinContext
14   rdf:resource="https://example.org/medieval#PneumaTheory"/>
15 <medieval:withinContext
16   rdf:resource="https://dbpedia.org/page/Ancient_Greek_medicine"/>
17 <medieval:hasModernCounterpart
18   rdf:resource="http://dbpedia.org/resource/Artery"/>
19 <time:hasTime>
20   <time:TemporalEntity
21     rdf:about="http://www.example.org/medieval#Artery_Time">
22     <time:hasBeginning>
23       <time:Instant rdf:about="http://www.example.org/
24         medieval#Artery_Time_Beginning">
25         <time:inDateTime rdf:resource="http://www.example.org/
26           medieval#Antiquity_Medicine"/>
27       </time:Instant>
28     </time:hasBeginning>
29     <time:hasEnd>
30       <time:Instant rdf:about="http://www.example.org/
31         medieval#Artery_Time_End">
32         <time:inXSDgYear rdf:datatype="http://www.w3.org/
33           2001/XMLSchema#gYear">1628
34       </time:inXSDgYear>
35     </time:Instant>
36   </time:hasEnd>
37 </time:TemporalEntity>
38 </time:hasTime>
39 </owl:Class>
40 ...
41 </rdf:RDF>

```

**Abb. 2** | Auszug aus der Ontologie mit der Klasse *Artery*. CC0.

OWL-Relation *inverseOf*<sup>71</sup> (Z. 16) wird eine semantische Beziehung eingeführt, die es ermöglicht, automatische Schlussfolgerungen über die Ontologieklassen und die mit ihnen verknüpften Entitäten der LD-Ressourcen (siehe Anm. 71) zu ziehen.<sup>72</sup>

<sup>71</sup> <https://www.w3.org/TR/owl-ref/#inverseOf-def> (Zugriff: 10.03.2025).

<sup>72</sup> Zum *Automatic Reasoning* vgl. Grimm, Abecker, Völker u. a. 2011, S. 515–519.

Die Definition des Konzepts, das durch den Begriff *Artery* repräsentiert wird, erfolgt über das Prädikat `rdfs:comment` und ein Literal (Abb. 2, Z. 4–7). Dieses Literal beschreibt, dass die ‚Arterie‘ in der mittelalterlichen Physiologie (geprägt von der antiken Pneumalehre, siehe oben) als eine Blutbahn begriffen wird, die, vom Herzen ausgehend, das ‚vitalisierende‘ Blut mit dem *zotikón* (griech.; lat. *spiritus vitalis*) in allen Körperteilen verteilt. Dieses Pneuma ist zuständig für die Regulierung der Körperwärme und die Aufrechterhaltung des Lebens. Der Blutkreislauf ist noch nicht entdeckt und das Konzept der ‚Arterie‘ beinhaltet, dass das ‚vitalisierende‘ Blut in den Arterien auch wieder zum Herzen zurückfließt. Der Begriff *Artery* ist eine Unterkategorie des Begriffs *BloodVessel* (Z. 9); sein Konzept steht im Erklärungskontext der Pneumalehre (Z. 10) und der antiken griechischen Medizin (Z. 11). Um Interoperabilität mit ‚Dbpedia‘ zu erreichen und eine Verknüpfung der ‚mittelalterlichen Arterie‘ mit der ‚modernen Arterie‘ zu ermöglichen, nutzt ALMA die Relation `hasModernCounterpart`, die auf den ‚Dbpedia‘-Eintrag *Artery* (modernes Konzept) verlinkt (Z. 12).

Für die Modellierung der ‚Lebenszeit‘ des Begriffskonzepts wird ein Zeitintervall (Abb. 2, Z. 13–27) per Begriff definiert. Es baut auf der jeweiligen Belegsituation der sprachlichen Zeichen auf, die das Begriffskonzept repräsentieren. Da ALMA sich empirisch auf vier Sprachen stützt und die Erkenntnisse durch weitere Sprachen (Mittellatein, Altspanisch und -katalanisch) absichert, ist eine gewisse übersprachliche Gültigkeit zu erhoffen. Die Angabe des Zeitintervalls kann mit Hilfe der ‚Time Ontology in OWL‘<sup>73</sup> erfolgen, die es ermöglicht, Zeitpunkte und Zeitintervalle zu definieren. Der Zeitraum wird modelliert als eine *TemporalEntity* der ‚Time Ontology in OWL‘ („A temporal interval or instant“<sup>74</sup>), deren Relationen `hasBeginning` (Z. 15–19) und `hasEnd` (Z. 20–25) auf je einen Instant („A temporal entity with zero extent or duration“<sup>75</sup>) als Anfangs- und Endpunkt verweisen und damit die Grenzen des Intervalls beschreiben. Wenn für ein Konzept die sprachlichen Realisierungen nur in genau einem Text gefunden werden, der Zeitraum also einem Zeitpunkt gleichkommt, erhalten Anfangs- und Endpunkt dasselbe Datum. Der Datentyp *gYear* (Z. 22) codiert ein Jahr im gregorianischen Kalender.<sup>76</sup> Die Tatsache, dass bei Konzept-/Bedeutungsentwicklungen die Grenzen zwischen einem gegebenen Konzept und einem sich daraus entwickelnden neuen Konzept nicht immer trennscharf sind, erfordert das Erproben eines feingranularen Auszeichnungsstandards, um zeitlich überlappende Übergangsphänomene zu erfassen.

73 <https://www.w3.org/TR/owl-time/> („published [15 November 2022] by the Spatial Data on the Web Working Group as a Candidate Recommendation Draft“) (Zugriff: 10.03.2025).

74 <https://www.w3.org/TR/owl-time/#time:TemporalEntity> (Zugriff: 10.03.2025).

75 <https://www.w3.org/TR/owl-time/#time:Instant> (Zugriff: 10.03.2025).

76 <https://www.w3.org/TR/xmlschema11-2/#gYear> (Zugriff: 10.03.2025).

Für die Angabe des Zeitintervalls von ‚Arterie‘ muss nun auf die Belegsituation geschaut werden: Das Konzept muss theoretisch mit Harveys Entdeckung des Blutkreislaufs von 1628 enden, die die Vorstellungen vom Blutgefäßsystem umfassend korrigierte. Solange die Letztbelege – mittels lexikalisch-semanticischer Analyse durch das Projekt ALMA – noch nicht im Detail recherchiert sind, ist hier provisorisch ‚1628‘ als Letztdatum anzunehmen. Erstmals belegt ist im Altfranzösischen die Bezeichnung *artere* mit der pneumatheoretischen Bedeutung in ‚Li fet des Romains ou Roman de Jules César‘ von 1213.<sup>77</sup> Die Bezeichnungen in den anderen Sprachen sind noch zu recherchieren. Allerdings geht das mittelalterliche Konzept auf die antike Medizin zurück, wo die Pneumalehre und die damit verbundene sehr alte Frage, ob die Seele als ein wenn auch noch so feiner Stoff zu verstehen ist, bereits von hippokratischen Ärzten diskutiert wird.<sup>78</sup> Als (provisorisches) Datum 1213 anzugeben, wäre also irreführend. Dass ein Konzept antike Vorstellungen tradiert und das Anfangsdatum des Zeitintervalls somit in der Zeit der antiken Medizin liegt, wird häufig der Fall sein. Eine Idee für diesen Fall ist es, immer auf genau eine Entität *Antiquity\_Medicine* zu verweisen, die als *DateTimeDescription* der ‚Time Ontology in OWL‘<sup>79</sup> definiert und mit Zeitintervall präzisiert werden muss. Damit ließen sich Redundanzen vermeiden.

Die sprachlichen Zeichen, die die Begriffskonzepte der Ontologie repräsentieren, zum Beispiel altfranzösisch *artaire*, mittelfranzösisch *arthere*, *aurtre*,<sup>80</sup> italienisch (14. Jahrhundert) *arteria*,<sup>81</sup> altokzitanisch *artaria*, *erceria*<sup>82</sup> etc., werden als Teil der lexikalisch-semanticischen Studien als LOD modelliert (siehe Anm. 71). In der Modellierung verweisen sie über die oben genannten Relationen der ‚OntoLex-Lemon‘-Ontologie auf die Klassen der historisierten Ontologie.<sup>83</sup>

## 5 Fazit

Das hier beschriebene lexikalisch-semanticische Mapping von *Linked Open Data* auf Begriffe von Domänenontologien ist ein sehr vielversprechender Weg, einen

---

<sup>77</sup> Li fet des Romains III, Kap. 1, § 14, Bd. 1, S. 362, Z. 12; vgl. Tittel 2004, S. 198.

<sup>78</sup> Vgl. Putscher 1974, S. 8–24, 42–55.

<sup>79</sup> <https://www.w3.org/TR/owl-time/#time:DateTimeDescription> (Zugriff: 10.03.2025).

<sup>80</sup> <https://deaf.hadw-bw.de/lemme/artere> (Zugriff: 10.03.2025; bald <https://deaf.ub.uni-heidelberg/lemme/artere>).

<sup>81</sup> Vgl. Calò u. Lupis 1989, Sp. 1473.

<sup>82</sup> Vgl. Deggeller 1997, S. 367.

<sup>83</sup> Vgl. Prifti, Selig, Schweickard u. a. 2023, S. 317, für einen Auszug aus der LOD/RDF-Modellierung der Beispielstudie zum ‚Bauchfell‘, sprachlich realisiert z. B. durch afrz. *mirac*. (Die *Domaine* <<https://alma.hadw.bw.de>> ist dort zu korrigieren in <<https://alma.hadw-bw.de>>.)

semantischen Zugang zu linguistischen Ressourcen wie Texten, Grammatiken und Wörterbüchern zu erhalten. Das Modellieren als LOD erzeugt mit Hilfe der gemeinsamen Formate und des Mappings Interoperabilität über Ressourcen, Sprachen, Kultur- und Zeiträume hinweg: Es standardisiert Ressourcen, ohne bestehende Datenstrukturen zu beeinträchtigen, es macht maschinenlesbar und abfragbar, was sonst in Datensilos einem offenen Zugriff verwehrt ist. Es kreiert eine Schnittstelle zwischen heterogenen historischen Inhalten und ermöglicht einen Datenzugriff, der dieser Heterogenität gewachsen ist.

Das Mapping ermöglicht komplexe Abfragen über die Dinge und ihre Relationen, die in der Ontologie abgebildet sind. Für historische Ressourcen sind dabei historisierte Domänenontologien maßgeblich, um Anachronismen im Mapping zu vermeiden, die den Datenzugriff und darauf aufbauende Forschungen verfälschen würden. Dabei sind die Ontologien in zweierlei Hinsicht flexibel und anschlussfähig an andere Disziplinen, Domänen, zeitliche oder kulturelle Fokusse: Sie sind zum einen erweiterbar mit Verknüpfungen mit Entitäten anderer Ontologien, zum anderen mit Klassen und Relationen, um neues Wissen der Domäne zu erschließen und für das Mapping bereitzustellen. Und das Erarbeiten von Ontologien ist weit mehr als Mittel zum Zweck: Erfahrungen zeigen, dass das Klassifizieren, Hierarchisieren und Vernetzen von Domänenwissen bei bislang darin unerfahrenen Domänenexperten Impulse für neue Perspektiven und neue Forschungsfragen zu generieren vermag.

## Literaturverzeichnis

### Quellen

Li fet des Romains, compilé ensemble de Saluste et de Suetoine et de Lucan. Hrsg.

v. Louis-Fernand Flutre u. Kornelis Sneijders de Vogel. 2 Bde. Paris 1935.

### Forschungsliteratur

#### **Baldinger, Kurt, Frankwalt Möhren u.**

**Thomas Städtler (Hgg.):** Dictionnaire étymologique de l'ancien français – DEAF (fondé par Kurt Baldinger, continué par Frankwalt Möhren, publié sous la direction de Thomas Städtler). Québec, Tübingen, Berlin 1971–2020. Elektronische Version DEAFplus / DEAFpré: <https://deaf.ub.uni-heidelberg.de>.

#### **Bellandi, Andrea, Emiliano Giovannetti,**

**Silvia Piccini u. a.:** Developing LexO. A Collaborative Editor of Multilingual Lexica and Terminology-Ontological Resources in the Humanities. In: Francesca Frontini, Larisa Grčić Simeunović, Špela Vintar u. a. (Hgg.): Proceedings of the Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS) 2017. Montpellier 2017, S. 11–21.

- Berners-Lee, Tim, James Hendler u. Ora Lassila:** The Semantic Web. In: *Scientific American* 284 (2001), S. 34–43.
- Bizer, Christian, Tom Heath u. Tim Berners-Lee:** Linked Data – The Story so Far. In: *International Journal on Semantic Web and Information Systems* 5 (2009), S. 1–22.
- Blank, Andreas:** Einführung in die lexikalische Semantik. Tübingen 2001.
- Bozzi, Andrea:** Un’ontologia per il DiTMAO (Dictionnaire des Termes Médico-botaniques de l’Ancien Occitan). In: David Trotter, Andrea Bozzi u. Cédric Fairon (Hgg.): *Actes du XXVII<sup>e</sup> Congrès international de linguistique et de philologie romanes* (Nancy, 15–20 juillet 2013). Section 16. Nancy 2016, S. 55–63.
- Calò, Fernando u. Antonio Lupis:** arteria ‚trachea‘. In: Max Pfister, Elton Prifti u. Wolfgang Schweickard (Hgg.): *Lessico Etimologico Italiano – LEI*, Bd. 3.1. Wiesbaden 1989, Sp. 1472–1477.
- Chiarcos, Christian, John McCrae, Philipp Cimiano u. a.:** Towards Open Data for Linguistics. *Lexical Linked Data*. In: Alessandro Oltramari, Piek Vossen u. Eduard Hovy (Hgg.): *New Trends of Research in Ontologies and Lexical Resources*. Heidelberg 2013, S. 7–25.
- Cimiano, Philipp, Christian Chiarcos, John P. McCrae u. a.:** *Linguistic Linked Data. Representation, Generation and Applications*. Cham 2020.
- Declerck, Thierry, Eveline Wandl-Vogt, Eveline u. Karlheinz Mörth:** Towards a Pan European Lexicography by Means of Linked (Open) Data. In: Iztok Kosem, Miloš Jakubiček, Jelena Kallas u. a. (Hgg.): *Electronic Lexicography in the 21st Century. Linking Lexical Data in the Digital Age. Proceedings of the eLex 2015 Conference*. Ljubljana, Brighton 2015, S. 342–355.
- Deggeller, Marie-José:** ‚arteria‘. In: Walther von Wartburg (Hg.): *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*, Bd. 25. Basel 1997, S. 367–369.
- Dittrich, Andreas:** Historische Ontologie als *intellectual history*? In: IASLonline [21.02.2006]. [https://www.iaslonline.lmu.de/index.php?vorgang\\_id=1635](https://www.iaslonline.lmu.de/index.php?vorgang_id=1635) (Zugriff: 18.05.2024).
- Färber, Michael, Frederic Bartscherer, Carsten Menne u. a.:** Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and Yago. In: *Semantic Web* 9 (2017), S. 1–53.
- Grimm, Stephan, Andreas Abecker, Johanna Völker u. a.:** Ontologies and the Semantic Web. In: John Domingue, Dieter Fensel u. James A. Hendler (Hgg.): *Handbook of Semantic Web Technologies*. Heidelberg 2011, S. 507–579.
- Guarino, Nicola, Daniel Oberle und Steffen Staab:** What is an Ontology? In: Steffen Staab u. Rudi Studer (Hgg.): *Handbook on Ontologies*. Berlin, Heidelberg 2009, S. 1–17.
- Hacking, Ian:** *Historical Ontology*. Cambridge, MA, London 2002.
- Hallig, Rudolf u. Walther von Wartburg:** Begriffssystem als Grundlage für die Lexikographie/Système raisonné des concepts pour servir de base à la lexicographie. 2. Aufl. Berlin 1963.
- Ide, Nancy u. David Woolner:** Historical Ontologies. In: Khurshid Ahmad, Christopher Brester u. Mark Stevenson (Hgg.): *Words and Intelligence II: Essays in Honor of Yorick Wilks*. Berlin, Heidelberg 2007, S. 137–152.
- Kay, Christian, Jane Roberts, Michael Samuels u. a. (Hgg.):** *Historical Thesaurus of the Oxford English Dictionary, with Additional Material*

- from *A Thesaurus of Old English*. 2 Bde. Oxford 2009.
- Khan, Anas, Christian Chiarcos, Thierry Declerck u. a.:** When Linguistics Meets Web Technologies. Recent Advances in Modelling Linguistic Linked Data. In: *Semantic Web* 13 (2022), S. 1–64.
- Lezcano, Leonardo, Salvador Sánchez-Alonso u. Antonio Roa-Valverde:** A Survey on the Exchange of Linguistic Resources. In: *Program* 47 (2013), S. 263–281.
- Lindén, Jan-Ivar:** Prolegomena zur historischen Ontologie. Heidelberg 2019.
- McCrae, John, Elena Montiel-Ponsoda u. Philipp Cimiano:** Integrating WordNet and Wiktionary with 'Lemon'. In: Christian Chiarcos (Hg.): *Linked Data in Linguistics*. Berlin, Heidelberg 2012, S. 25–34.
- Meroño-Peñuela, Albert, Ashkan Ashkpour, Marieke Erp u. a.:** Semantic Technologies for Historical Research: A Survey. In: *Semantic Web* 6 (2014), S. 539–564.
- Nannini, Alessandro Alfredo:** La mappatura semantica del Lessico Etimologico Italiano (LEI). In: Lidia Becker, Julia Kuhn, Christina Ossenkop u. a. (Hgg.): *Digitale romanistische Sprachwissenschaft. Stand und Perspektiven (Romanistisches Kolloquium 34)*. Tübingen 2023, S. 119–145.
- Pease, Adam u. Ian Niles:** Toward a Standard Upper Ontology. In: Chris Welty u. Barry Smith: *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*. New York 2001, S. 2–9.
- Pfister, Max, Elton Prifti u. Wolfgang Schweickard (Hgg.):** *Lessico Etimologico Italiano – LEI*. Wiesbaden 1979–.
- Prifti, Elton, Wolfgang Schweickard, Maria Selig u. a.:** Sprachdatenbasierte Modellierung von Wissensnetzen in der mittelalterlichen Romania (ALMA): Projektskizze. In: *Zeitschrift für romanische Philologie* 139 (2023), S. 301–322.
- Putscher, Marielene:** *Pneuma, Spiritus, Geist. Vorstellungen vom Lebensantrieb in ihren geschichtlichen Wandlungen*. Wiesbaden 1974.
- Roberts, Jane A., Christian Kay u. Lynne Grundy:** *A Thesaurus of Old English*. 2. Aufl. Amsterdam 2000.
- Romein, C. Annemieke, Max Kemman, Julie Birkholz u. a.:** State of the Field: Digital History. In: *History* 105/365 (2020), S. 291–312.
- Schipperges, Heinrich:** *Geschichte der Medizin in Schlaglichtern*. Mannheim 1990.
- Tittel, Sabine:** *Die Anatomie in der Grande Chirurgie des Gui de Chauliac*. Tübingen 2004.
- Tittel, Sabine:** Kulturerbe, historische Linguistik und Semantic Web: Eine Einführung mit Fallbeispiel zu französischen linguistischen Ressourcen. In: Lidia Becker, Julia Kuhn, Christina Ossenkop u. a. (Hgg.): *Digitale romanistische Sprachwissenschaft. Stand und Perspektiven (Romanistisches Kolloquium 34)*. Tübingen 2023, S. 149–191.
- Tittel, Sabine:** Historische lexikalische Semantik und Linked Data. Modellierung von Ressourcen der mittelalterlichen Galloromania für das Semantic Web. Berlin, Boston 2024.
- Tittel, Sabine u. Christian Chiarcos:** Historical Lexicography of Old French and Linked Open Data: Transforming the Resources of the *Dictionnaire étymologique de l'ancien français* with OntoLex-Lemon. In: Nicoletta Calzolari, Kahlid Choukri, Christopher Cieri u. a. (Hgg.): *Proceedings of the*

Eleventh International Conference on Language Resources and Evaluation (LREC 2018). GLOBALEX Workshop (GLOBALEX-2018). Paris 2018, S. 58–66.

**Tittel, Sabine, Frances Gillis-Webber u.**

**Alessandro A. Nannini:** Towards an Ontology Based on Hallig-Wartburg's Begriffssystem for Historical Linguistic Linked Data. In: Maxim Ionov, John P. McCrae, Christian Chiarcos u. a. (Hgg.): Proceedings of the International Conference on Language Resources and

Evaluation (LREC 2020). 7th Workshop on Linked Data in Linguistics (LDL-2020). Marseille 2020, S. 1–10.

**Wang, Shenghui, Stefan Schlobach u.**

**Michel C. A. Klein:** Concept Drift and How to Identify it. In: Journal of Web Semantics. Science, Services and Agents on the World Wide Web 9 (2011), S. 247–265.

**Weyl, Hermann:** Philosophy of Mathematics and Natural Science. Princeton 1949.

# Linked Open Data and Manuscript Descriptions

## Contact

### Bernhard Bauer,

Universität Graz, Institut für Digitale Geisteswissenschaften,  
Elisabethstraße 59, 3. Stock,  
8010 Graz, Österreich,  
bernhard.bauer@uni-graz.at  
 <https://orcid.org/0000-0003-2881-0972>

### Roman Bleier,

Universität Graz, Institut für Digitale Geisteswissenschaften,  
Elisabethstraße 59, 3. Stock,  
8010 Graz, Österreich,  
roman.bleier@uni-graz.at  
 <https://orcid.org/0000-0003-4674-1042>

### Jakob Sonnberger,

Universität Graz, Institut für Digitale Geisteswissenschaften,  
Elisabethstraße 59, 3. Stock,  
8010 Graz, Österreich,  
jakob.sonnberger@uni-graz.at

**Abstract** This chapter explores the intersection of Linked Open Data (LOD) and manuscript descriptions, emphasising the importance of data availability under the LOD principles for research in the humanities. It begins by outlining the principles of Linked Data, including Tim BERNERS-LEE's five-star LOD model and technologies such as RDF that underpin the Semantic Web. The discussion then shifts to the application of these technologies to medieval studies, highlighting how archives and libraries are making resources available online. The article delves into the Text Encoding Initiative (TEI) and its role in creating machine-readable manuscript descriptions, evaluating selected digital scholarly editions for their use of LOD. Finally, a case study on the GAMS repository at the University of Graz demonstrates current practices and proposes a possible approach to unifying manuscript ontology to enhance interoperability and scholarly utility in a real-world example. The study critically discusses the current status of LOD for medieval manuscript studies and underscores its potential to improve research accessibility and connectivity in the future.

**Keywords** Digital Medieval Studies; Linked Open Data; Manuscript Studies; Ontologies; Semantic Web

## 1 Introduction

The practice of describing medieval manuscripts, a process refined over centuries, varies greatly across regions, institutions, and intellectual traditions. Manuscript descriptions also serve different purposes: they are specifically created for manuscript catalogues, scholarly articles, or scholarly editions, differing in structure and detail. All adhere to a common set of essential requirements, highlighting their shared intent.<sup>1</sup> Nevertheless, their varied forms also cause issues, as they make interoperability, an essential aspect of research data, difficult across catalogues or manuscript databases. Since the early days of the World Wide Web (WWW), the attempts to create overarching databases of manuscript descriptions have only been partially successful.<sup>2</sup> However, we are still a long way from a global manuscript database. Nevertheless, if research and institutions continue to share their manuscript descriptions following the Linked Open Data (LOD) principles, this will bring us closer to reaching this goal in the near future.

The relevance of data for research, even in the humanities, has been strongly emphasised by recent developments in machine learning and Artificial Intelligence (AI), which have a huge impact on basically all aspects of society, including our research. Therefore, the recommendation by Tim BERNERS-LEE, the inventor of the WWW, to publish data on the web<sup>3</sup> emphasises the importance of openness, reusability, and meaningful connections between data. The result is a web of data, often referred to as the Semantic Web, which, ideally, can be queried like a global database.<sup>4</sup> The reality, however, is that many data are still trapped in self-contained data silos, and to make them interoperable is a big and expensive task. Manuscript descriptions stored in institutional databases are no exception. The cause might be proprietary data formats, different data models, or manuscript descriptions of different levels of detail. Linking such datasets together using LOD would be a big step forward, creating new opportunities for research. The Mapping Manuscript Migrations project explored such a scenario. Data about pre-modern manuscripts from three databases were interlinked using LOD technologies to explore the movement of these manuscripts. The project has shown the potential

---

1 Beadle and Hanna 2020.

2 For instance, projects such as ‘Manuscriptorium’, a sub-aggregator of ‘Europeana’ for historical books, provide a virtual research environment or digital library around the collected data. Other examples are ‘Manuscripta’ and ‘Mediaevalia’, recently replaced by the ‘Handschriftenportal’, for manuscript data from German collections; ‘Bibliissima’ for collections in France; or the virtual manuscript library ‘E-Codices’ in Switzerland. Weidenmann 2020.

3 Berners-Lee 2006.

4 To get an idea of the amount of available LOD, the LOD Cloud is a regularly updated website showing the biggest data sets, their domain, and connections as a diagram: <https://lod-cloud.net/> (accessed: 25/11/2024).

of LOD, but also the limitations caused by data inconsistencies in the three very different datasets.<sup>5</sup>

In our chapter, we will explore general concepts of LOD for the description of manuscripts and also take a critical look at the limitations and issues that arise. The recommendations of the Text Encoding Initiative (TEI) for encoding manuscript descriptions and preparing data for the Semantic Web are important, as they are used by libraries, archives, and digital scholarly editions. As our background is in digital scholarly editing, the final part of the paper will focus on the implementation of LOD in this context.<sup>6</sup> We look at a real-life use case and propose a possible approach for the manuscript projects currently carried out at the Department of Digital Humanities (DDH) at the University of Graz.

## 2 LOD

### 2.1 From Data to LOD

In its first years, the WWW was primarily used to publish documents online and to create connections in the form of hyperlinks between these documents or webpages. One webpage would make a connection to another webpage and, thus, create a web of documents. These documents contained data on different things and topics, but in a format (presentational markup) that was intended to be read and interpreted by humans. Also, manuscript descriptions were published in this way on the WWW. A project-specific data model may cause interoperability issues with other databases. Furthermore, a restrictive licence prevents the automated processing and reuse of the data. LOD and the Semantic Web are often considered to be the solution.<sup>7</sup>

BERNERS-LEE described a five-step ranking system for open data: five-star LOD.<sup>8</sup> One star is achieved if one makes documents without structure or metadata available under an open licence, for instance, an image of a table. A human viewer can probably read the image correctly as a table, but the tabular data cannot be processed correctly by a program. Two stars are reached if the data are provided both under an open licence and in a structured format, for instance, an Excel table under a CC BY licence. If the same structured data are provided in a non-proprietary open format, this qualifies as three-stars open data. The next step would be

---

<sup>5</sup> Burrows 2022.

<sup>6</sup> See, for instance, Witt 2016; Vogeler 2019.

<sup>7</sup> See, for instance, Heath and Bizer 2011.

<sup>8</sup> Berners-Lee 2006.

to use the Resource Description Framework (RDF) standard for description and to provide URIs, so others can point to your data and make connections. Finally, five-star LOD is reached if one links data with other data to provide context. This might be a reference to another database or an authority file.

RDF is a key standard to reach the five-star level. RDF is organised in statements that are represented as RDF triples, sometimes also semantic triples. These RDF triples describe data and relationships between data as subject–predicate–object (S-P-O) relationships, similar to simple sentences. As an example, we may state that “Book of Armagh”(S) “is held by”(P) the “Library of Trinity College Dublin”(O). In most cases, these literal statements can also be translated into a representation using URIs instead of literal strings. Many such connected triples form a network. Eventually, this enables machines to understand and interpret the information in a more sophisticated and meaningful way. Ontologies play a key role in this process of interlinking data, as they are a formal definition of concepts and relationships that is used to link individual data.

While a data model is a representation of how data is structured and organised in a database, ontologies, as used in the context of the Semantic Web and Digital Humanities, are a representation of knowledge by defining the relationships between different concepts. The primary purpose<sup>9</sup> of developing an ontology is to enhance the understanding of a specific domain (domain ontology), such as archives or museum collections. It serves as the foundation for developing a well-functioning data model, which can be implemented as a database compatible with the ontology, providing a concrete model for conceptualising the domain. An upper or top-level ontology, on the other hand, provides general concepts and functions as an ontology across various more specific domain ontologies.<sup>10</sup>

## 2.2 Authority Files and IIF

For identifying and addressing ‘things’ on the WWW, so-called URIs (Uniform Resource Identifiers) and URLs (Uniform Resource Locators) have been used since the early 1990s.<sup>11</sup> In the context of the Semantic Web, URIs are used as globally unique digital identifiers. Having a URI as an identifier for something (e. g. manuscript, person) allows that other data can be linked to it and RDF statements can be made about it and, hence, more data can be associated. For instance, the ‘Book

---

<sup>9</sup> Eide and Ore 2018, p. 184.

<sup>10</sup> Rehbein 2017 and Guarino 1998 suggest distinguishing top-level-ontology, domain ontology, task ontology, and application ontology.

<sup>11</sup> Berners-Lee 1994.

of Armagh' example from before could also be represented using 'Wikidata' URIs: `wd:Q997846 - wdt:P195 - wd:Q856559`.<sup>12</sup>

Authority files, or authority records, are databases often maintained by libraries, museums, or archives that provide data about authors, subjects, or places. An entry in an authority file also has an URI for identification and referencing within the WWW. The use of authority file URIs allows one to identify data about the same individual, for instance, an author, subjects, or places in different datasets and databases. The GND ('*Gemeinsame Normdatei*') is the authority file maintained by the German National Library based on its library records and data, and there are many other libraries and institutions across the world providing valuable authority records. Data in the GND, TGN, and other institutional authority files is usually curated by the publishing institution, and updates or changes to entries by users are not intended. 'Wikidata', on the other hand, is a project by the Wikimedia Foundation and a collaborative authority file. Similarly to 'Wikipedia', users can create and edit entries. On the one hand, the benefit is that far more terms are covered in 'Wikidata', but the downside is that user-generated data is usually not very consistent and of varying quality.<sup>13</sup>

The Mapping Manuscript Migrations project has done substantial research on the use of authority files in manuscript description databases and concluded that "generalist vocabularies like VIAF and TGN already have a reasonably good coverage of persons, places, and organisations".<sup>14</sup> TGN ('*Getty Thesaurus of Geographic Names*') is an authority file provided by the Getty Research Institute, which also provides several others. VIAF ('*Virtual International Authority File*'), on the other hand, functions as a joint international authority file, and its entries contain references to other authority files.

An international authority file for manuscripts is still a desideratum, but the need to create such a database has been discussed for years. The International Standard Manuscript Identifier (ISMI)<sup>15</sup> was such an attempt but has been dormant for some years. In the German-speaking regions, the platform '*Manuscripta Mediaevalia*', a digital catalogue developed since 1996, provided identifiers and metadata on manuscripts in German institutions. The project was recently replaced by the

---

<sup>12</sup> The namespace `wd` stands for '<http://www.wikidata.org/wiki/>' and `wdt` for '<http://www.wikidata.org/wiki/Property:>'.

<sup>13</sup> A recent evaluation of the use of 'Wikidata' in DH projects reports on the wide use of 'Wikidata', but also on the concerns about data quality and bias. "The quality of the data is influenced by the general practices, tactics, and strategies of other players in the specific domain." Zhao 2023.

<sup>14</sup> Burrows 2020, p. 217.

<sup>15</sup> *Ibid.*, p. 216.

‘Handschriftenportal’,<sup>16</sup> which brings the conceptual idea of an online catalogue for manuscripts to the next level and aims to become an authority file for manuscripts. The ‘Handschriftenportal’ already provides LOD about the manuscripts in the database and plans to implement an application programming interface (API) to query the data in its second phase.

The ‘Handschriftenportal’ also uses another important technology for digital manuscript studies, the IIF (International Image Interoperability Framework), a specification for APIs developed by libraries, museums, and other organisations to provide access to images and related data. Using the IIF, manuscript images can be shared by an institution in an interoperable manner, and online projects, e. g. digital editions or collections, can easily reuse the manuscript images and include them on their websites. The IIF-manifest, a document written in yet another Semantic Web standard, JSON-LD (JavaScript Object Notation for Linked Data), contains all the information for displaying the manuscript images to a user, such as the structure of the manuscript, image links, essential metadata, and (copyright) conditions for use. Additionally, IIF URIs are built in such a way that not only entire images can be addressed but also sections within an image. This is especially advantageous for annotating images or if one only wants to present a specific part of a manuscript page – for instance a gloss or marginalia. Digital scholarly editing also benefits from the application of IIF in libraries and archives, as it facilitates sharing manuscript descriptions and images as well as the development of resource-rich editions.

### 3 Manuscript Description and the Semantic Web

#### 3.1 TEI

The idea of making documents not only accessible to humans but also machine-readable and enriching them with semantic markup and structured metadata is one that humanities scholars have been doing and discussing since before the invention of the WWW. The most prominent example is TEI, which started in the 1980s and continuously maintains a standard for semantically enriching humanities data. From the beginning, scholarly editing and, consequently, the description of manuscripts was relevant to the TEI.<sup>17</sup> The TEI’s msdescription module offers a broad set of possibilities for medievalists to describe their objects of research

---

<sup>16</sup> Handschriftenportal 2018-, especially <https://handschriftenportal.de/info/services> (accessed: 25/11/2024). See also Weidenmann 2020.

<sup>17</sup> The ‘European MASTER project’ greatly influenced the development of the recommendations for manuscript description from the late 1990s. Burnard 2013.

semantically.<sup>18</sup> However, the TEI was created a good decade before Berners-Lee's vision of the Semantic Web, at a time when digital editions were produced on CD-ROM media and the WWW and web editions were still somewhat uncharted territory. Therefore, TEI was not created with linked data in mind. More recently, elements and attributes have been added to provide possibilities to connect to authority files and to store LOD. For instance, in 2020, the `tei:standOff` element was introduced, which serves as a container element for linked data and other contextual information, and it is recommended to use it to store lists with references to authority files. A frequent use case in digital editing with the TEI is the linking of a place or person to an authority file, for which the TEI provides different mechanisms, such as the `tei:idno` element or the attribute `@ref` with the URI of the relevant authority file entry.

TEI is frequently used in archives and libraries as a standard to provide manuscript descriptions.<sup>19</sup> It is also a core standard in manuscript databases such as the 'Handschriftenportal' or 'Biblissima',<sup>20</sup> which we have briefly mentioned above, and the main standard for digital editing, which is also our use case in this chapter.

### 3.2 Ontologies for Manuscript Studies

Since the turn of the century, more and more projects have begun to model the reality of manuscripts and their texts in specifically tailored ontologies. In general, we can talk about two different approaches that ontologies take in the Digital Humanities: *document-centric* and *event-centric*. Most of the ontologies established for digital editions so far employ the latter approach and usually reuse global models such as the CIDOC Conceptual Reference Model (CIDOC-CRM)<sup>21</sup> or the Simple Knowledge Organization System (SKOS).<sup>22</sup> A good overview of ontologies used to model data connected with antique and medieval manuscripts is presented by Sebastian BARZAGHI, Monica PALMIRANI, and Silvio PERONI.<sup>23</sup> They only present a single example for a document-centric approach, i. e. the 'Digitised Manuscripts to Europeana' project.<sup>24</sup> However, they mention several examples for ontologies taking an event-centric approach: the 'Henry III Fine Rolls' project,<sup>25</sup>

---

<sup>18</sup> Burghart 2017.

<sup>19</sup> Bellotto 2020.

<sup>20</sup> Schaßan 2022; Weidenmann 2020.

<sup>21</sup> [https://www.cidoc-crm.org/html/cidoc\\_crm\\_v7.1.3.html](https://www.cidoc-crm.org/html/cidoc_crm_v7.1.3.html) (accessed: 25/11/2024).

<sup>22</sup> <https://www.w3.org/2009/08/skos-reference/skos.html> (accessed: 25/11/2024).

<sup>23</sup> Barzaghi, Palmirani, and Peroni 2020, pp. 120–124.

<sup>24</sup> Dröge, Iwanowa, and Hennicke 2014.

<sup>25</sup> Vieira and Ciula 2007.

the ‘Sharing Ancient Wisdoms’ project,<sup>26</sup> ‘Biblissima’,<sup>27</sup> ZHITOMIRSKY-GEFFET and PREBOR’s<sup>28</sup> model for historical handwritten Hebrew manuscripts. Based on their review of these previous ontologies, BARZAGHI, PALMIRANI, and PERONI designed their own model to represent the resources found in the digital catalogue of ‘Progetto IRNERIO’.<sup>29</sup> This catalogue is a collection of the medieval texts on Roman and canon law found in the Royal College of Spain in Bologna. Their work resulted in an ontology which they called MeMO (‘Medieval Manuscript Ontology’).<sup>30</sup> The data model should enable the “preservation of the flow historical narrative [...] the representation of the context of the manuscripts [...] [and] the description of the structure of the material”.<sup>31</sup> They designed MeMO around FRBR (Functional Requirements for Bibliographic Records). This methodology employs an iterative workflow of small steps to create a well-developed and documented model. BARZAGHI, PALMIRANI, and PERONI concluded that “MeMO is sufficient to meet the domain expert’s requirements”.<sup>32</sup> But as far as we can see, the ontology has not been applied to the IRNERIO catalogue itself – or, at least, not in a publicly accessible way. Parts of MeMO are, however, reused in ‘GlossViBe’,<sup>33</sup> a project published in GAMS, the ‘Geisteswissenschaftliches (Humanities) Asset Management System’ of the University of Graz. Three other projects, currently in development in GAMS, are also basing their manuscript descriptions on this ontology, but the projects are not online yet: the DFG-FWF-funded ‘History as a Visual Concept: Peter of Poitiers’ Compendium Historiae’ (CHIGC), the ERC Starting Grant ‘German Arithmetical Treatises in Manuscripts of the Late Middle Ages (1400–1522)’ (ARITHMETIC), and the ERC Consolidator Grant ‘Celtic and Latin Glossing Traditions: Uncovering Early Medieval Language Contact and Knowledge Transfer’ (GlossIT).

### 3.3 Linked Open Data and Medieval Editions

So much for the theory; how many digital editions are actually using the Semantic Web technologies, and how much information do they convey about their manuscript sources? To explore this, we have taken a closer look at the editions

---

26 Jordanous, Hedges, Lawrence et al. 2012.

27 Gehrke, Frunzeanu, Charbonnier et al. 2015.

28 Zhitomirsky-Geffet and Prebor 2016.

29 <http://irnerio.cirsfid.unibo.it/> (accessed: 25/11/2024).

30 For a comprehensive introduction to MeMO, see Barzaghi, Palmirani, and Peroni 2020.

31 *Ibid.*, p. 120.

32 *Ibid.*, p. 136.

33 <https://gams.uni-graz.at/glossvibe> (accessed: 25/11/2024).

mentioned in Patrick SAHLE's catalogue of digital editions.<sup>34</sup> Building on this, we have established a spreadsheet including information from 140 digital editions<sup>35</sup> on titles, publication dates, periods covered (early, high, or late medieval), and whether or not their source files are available. Furthermore, we have recorded which editions offer RDF, NORMDATA, GND, VIAF, 'Wikidata', and 'Wikipedia' links.

The collected data spans 2006–2023, and we found that there is a discernible trend towards greater adoption of LOD principles over time. Nevertheless, only slightly more than a quarter of the 140 editions share their TEI source files in such a way that they can be immediately found.<sup>36</sup> Most of them (i. e. 38) are editions that were published after 2010. As TEI is a non-proprietary and open community standard, these editions would achieve at least three stars in the five-star LOD model. An additional step towards five-star LOD would include references to external data sets, especially authority files. Early editions rarely link to external data sets. However, more recent works demonstrate a marked improvement in this regard, reflecting the growing recognition of the value of interoperability and enriched metadata in the Digital Humanities. From 2013 onwards, editions such as 'Queste del saint Graal' (2013)<sup>37</sup> and Hartmann von Aue: 'Der arme Heinrich' (2018)<sup>38</sup> begin to show occasional links to authority files and databases (for instance, VIAF), but these are sporadic. More recent editions exhibit a higher tendency to incorporate linked data elements. For instance, 'Documenta Nepalica – Editions' (2021)<sup>39</sup> and 'CoReMA: Cooking Recipes of the Middle Ages' (2020)<sup>40</sup> link to multiple datasets such as GND, VIAF, 'Wikidata', and 'Wikipedia'. The editions from 2023, 'Tabula Salomonis'<sup>41</sup> and 'GlossViBe',<sup>42</sup> show some incorporation but still have room for more comprehensive linkage. This shift underscores the

---

34 Sahle, Vogeler, Klinger et al. 2020 ff. Sahle's catalogue currently (last change 6/6/2024) has 846 editions listed. We worked on a subset labelled as editions with a medieval focus. We are aware that Sahle's catalogue has certain biases (e. g. a Central European context) and is not a systematic collection of all digital editions worldwide. Nevertheless, it is still the most comprehensive catalogue that also includes early editions.

35 Of the 165 editions listed in the catalogue under the labels early, high, or late medieval, twenty-five could not be accessed during our research.

36 All three authors went through the list of editions and tried to find the source files. We consider source files that are so well hidden that none of the three of us could find them as not available.

37 <http://txm.ish-lyon.cnrs.fr/bfm/> (accessed: 25/11/2024).

38 <http://digi.ub.uni-heidelberg.de/ahd/> (accessed: 25/11/2024).

39 <https://nepalica.hadw-bw.de/nepal/editions> (accessed: 25/11/2024).

40 <https://gams.uni-graz.at/context:corema> (accessed: 25/11/2024).

41 <https://tabula.libripendis.eu/> (accessed: 25/11/2024).

42 <https://gams.uni-graz.at/glossvibe> (accessed: 25/11/2024).

importance of continuing to advance the integration of LOD in digital projects that use medieval manuscripts as their sources to enhance accessibility, discoverability, and scholarly utility. However, so far, only a few editions provide RDF in some form, and a fair share of those which do is published in GAMS.

## 4 A Case Study on Manuscripts in GAMS

### 4.1 Introducing GAMS

In the final part, we want to take a closer look at the practices employed by digital editions published in GAMS. GAMS is an infrastructure for the long-term preservation and publication of research data from humanities projects developed and maintained by the Department of Digital Humanities at the University of Graz.<sup>43</sup> GAMS shares essential metadata about its digital objects with aggregator projects such as ‘Europeana’ and ‘OpenAIRE’. However, there are only limited amounts of data shared from the manuscript descriptions. Nonetheless, individual GAMS projects can decide to provide more LOD within their projects. The preferred data format for long-term preservation of edition data is TEI. As a database, the repository uses a triplestore, and the standard workflow of a GAMS edition is to ingest TEIs and use an XSLT transformation to generate RDF (based on a project ontology) for the triplestore. While the triplestore is only used for internal queries, the RDF is accessible via the GAMS API and can be published in a digital edition.

As of November 2024, GAMS hosts more than eighty projects covering topics from a variety of epochs (from antiquity to the modern age) and geographical diversity (Graz and Styria, Austria; and international). For the present study, we concentrate on GAMS projects dealing with the Middle Ages. Filtering the list of published projects<sup>44</sup> yields eighteen projects falling into this period. Their topics range from art history and music (‘Aldersbach Chronicle’, ‘Cantus Network’) to cooking recipes (‘Cooking Recipes of the Middle Ages’) and health (‘Mittelalterlabor’) to legal topics (‘Database Supreme Jurisdiction’, ‘Legal Iconographic Database’), medieval works on biblical history (‘CHIGC’), museology (‘Virtual Museum of the University of Graz’), collections of seals and cartularies (‘Collection of Seals’, ‘Cartulary of the Duchy of Styria’), and early medieval Ireland (‘St Patrick’s epistles’, ‘Gloss-ViBe’, ‘GlossIT’). We can further narrow down the results and select only those projects that fit the topic of our chapter because their sources are medieval manuscripts and they make use of LOD and have already

---

<sup>43</sup> <https://gams.uni-graz.at/> (accessed: 25/11/2024).

<sup>44</sup> <https://gams.uni-graz.at/context:gams.projekte?mode=&locale=en> (accessed: 25/11/2024).

been published on GAMS (other projects are currently in development, but not yet on GAMS). Four projects remain: 1) ‘Cantus Network’,<sup>45</sup> which researches liturgy and music in the medieval ecclesiastical province of Salzburg; 2) ‘CoReMa’<sup>46</sup> recorded and analysed medieval cooking recipes found in 58 manuscripts; 3) ‘DEPCHA’,<sup>47</sup> a digital edition publishing cooperatively for historical accounts; and 4) ‘GlossViBe’,<sup>48</sup> a digital edition of the main text and the paratext of the manuscript Vienna, Österreichische Nationalbibliothek, Codex 15298 (Olim Suppl. 2698). In general, projects in the GAMS record quite a lot of data about the manuscripts in the TEI using elements from the msdescription-module, however, only very few are eventually available in the RDF.

## 4.2 Manuscript Ontology for GAMS

Most GAMS projects have individual data models that were developed based on project-specific requirements. This makes it difficult to query data across these projects. A common ontology for manuscript descriptions in GAMS projects will thus benefit both the repository and research and would offer multiple advantages. For example, it would enable researchers to quickly identify all respective manuscripts related to Middle High German texts or those dating from between 800 and 900 and facilitate cross-edition queries. Moreover, a common and pre-defined ontology would streamline the work on GAMS projects by avoiding the need to start from scratch for each new project. The resulting manuscripts could then be examined side-by-side, and cross-edition visualisations could be established. What should be recorded in a GAMS manuscript ontology, and do we risk running into any problems when establishing one? One thing that immediately comes to mind when researching medieval manuscripts and their texts is the issue of dating. Quite frequently, the manuscript dates do not conform with the text dates. As an example, we want to look at medieval Celtic literature. Apart from the early medieval glosses, most early texts are only transmitted in later manuscripts. Hence, the original date of the text is far from that of the manuscript in which it is recorded. An Irish example is, for instance, ‘Poems of Blathmac mac Con Brettan maic Conguso do Feraib Rois’. On account of linguistic, philological, and historical analyses, these two religious poems addressed to Mary, the mother of Christ,

---

<sup>45</sup> <https://gams.uni-graz.at/context:cantus> (accessed: 25/11/2024); Steiner and Klugseder 2019.

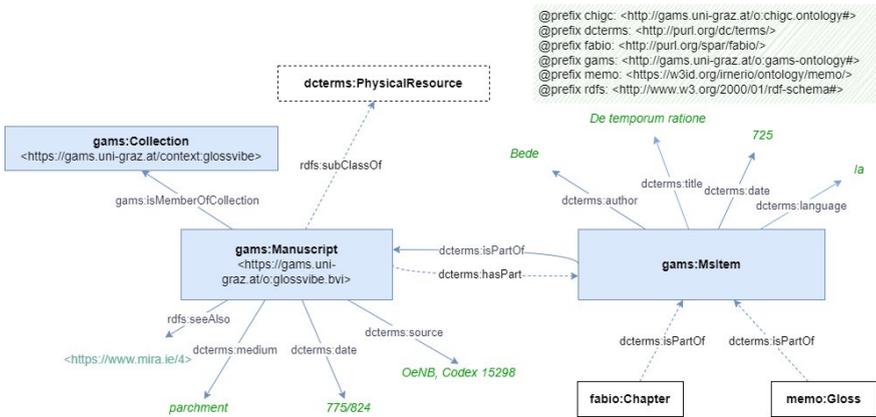
<sup>46</sup> [hdl.handle.net/11471/562.10](https://hdl.handle.net/11471/562.10) (accessed: 25/11/2024); Böhm, Eibinger, Klug et al. 2023.

<sup>47</sup> <https://gams.uni-graz.at/context:depcha> (accessed: 25/11/2024).

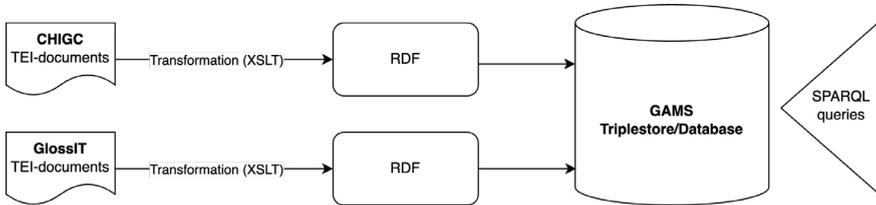
<sup>48</sup> See also Bauer 2024.

can be dated<sup>49</sup> to the Old Irish period. They are, however, only transmitted in a single manuscript witness dating from the seventeenth century. Simple queries on the manuscript date would therefore embed the text in completely the wrong era. Hence, our manuscript ontology should feature two different dates, one for the manuscript and one for the text.

Figure 1 illustrates our suggested data model for manuscript-based projects on the GAMS using ‘GlossViBe’ as an example. There are two classes for each manuscript: `gams:Manuscript` and `gams:MsItem`. The former records data in connection with the physical object, i. e. the manuscript itself, and the latter concerns an individual work as presented in the manuscript. The two classes are connected via a `hasPart/isPartOf` property. For the manuscript proper, we list information on its name (`dcterms:source`), date (`dcterms:date`), materiality (`dcterms:medium`), and a link to online manuscript catalogues (`rdfs:seeAlso`).



**Fig. 1 |** Proposed data model/ontology for the manuscript metadata in GAMS projects. Figure: Bauer, Bleier, and Sonnberger. CC-BY.



**Fig. 2 |** Using the example of CHIGC and GlossIT, the figure shows a simplified GAMS workflow to import data from the TEI documents to the triplestore. Figure: Bauer, Bleier, and Sonnberger. CC-BY.

49 See Stifter 2015.

The work part files the name of the author, the title, the date of the text, and the language (all `dcterms`). The model does not aim to offer exhaustive manuscript descriptions but is designed to facilitate cross-project searches for manuscript-based project data recorded in the triplestore (Fig. 2).

## 5 Conclusion

The integration of LOD with manuscript descriptions offers a transformative potential for medieval studies and the broader field of humanities research. The implementation of LOD standards facilitates a web of interconnected, machine-readable data, enhancing the accessibility and utility of scholarly resources. The chapter discusses LOD and ontologies and their application with a focus on manuscript descriptions for research and digital scholarly editing. Our exploration of various digital editing projects, particularly those hosted on GAMS, reveals a progressive trend towards adopting LOD principles. However, there remains significant scope for improvement, especially concerning ontologies for manuscript description, because although projects like MeMO have developed a fine-grained theoretical backbone, such ontologies are rarely applied in practice. For our own use case, we deem a less detailed ontology – which can be easily refined and expanded – more practical.

The proposed manuscript ontology for GAMS projects aims to standardise both descriptions and the generated RDF. This will, firstly, benefit operations with the GAMS triplestore; but as we build on existing ontologies, our work will, secondly, result in better RDF data that we provide. As we move forward in its development, the challenge lies in balancing the complexity and specificity of manuscript data with the need for interoperable, reusable, and widely accessible digital resources. The ongoing projects at the DDH at the University of Graz exemplify this endeavour, highlighting the importance of a unified approach to manuscript descriptions and a shared ontology. By continuing to refine and implement LOD frameworks, we can unlock new dimensions of scholarly research and ensure that the rich cultural heritage encapsulated in medieval manuscripts is preserved and made accessible for future generations.<sup>50</sup>

---

<sup>50</sup> This work is supported by the FWF grant ‘History as a visual concept’ (DOI: 10.55776/16133) and by the ERC grant ‘GlossIT’, Grant Agreement Number 101123203. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

## Bibliography

- Baierer, Konstantin, Evelyn Dröge, Kai Eckert et al.:** DM2E: A Linked Data Source of Digitised Manuscripts for the Digital Humanities. In: *Semantic Web 8* (2017), pp. 733–745.
- Barzaghi, Sebastian, Monica Palmirani, and Silvio Peroni:** Development of an Ontology for Modelling Medieval Manuscripts: The Case of Progetto IRNERIO. In: *Umanistica Digitale 9* (2020), pp. 117–140. <https://doi.org/10.6092/issn.2532-8816/11187> (accessed: 25/11/2024).
- Bauer, Bernhard:** Early Medieval Vernacular Celtic Glosses: Originals or Translations? A Case Study on the Vienna Bede [Version 2. Peer review: 1 approved; 2 approved with reservations]. In: *Open Res Europe 3/108* (2024). <https://doi.org/10.12688/openreseurope.16006.2> (accessed: 25/11/2024).
- Beadle, Richard and Ralph Hanna:** Describing and Cataloguing Medieval English Manuscripts: A Checklist. In: Orietta Da Rold and Elaine Treharne (eds.): *The Cambridge Companion to Medieval British Manuscripts* (Cambridge Companions to Literature). Cambridge 2020, pp. 13–38.
- Bellotto, Anna:** Medieval Manuscript Descriptions and the Semantic Web: Analysing the Impact of CIDOC CRM on Italian Codicological-Paleographical Data. In: *Digital Humanities Quarterly 14/1* (2020). <https://www.digitalhumanities.org/dhq/vol/14/1/000449/000449.html> (accessed: 25/11/2024).
- Berners-Lee, Tim:** Linked Data. 2006. <https://www.w3.org/DesignIssues/LinkedData.html> (accessed: 25/11/2024).
- Berners-Lee, Tim:** Universal Resource Identifiers in WWW. June 1994. <https://datatracker.ietf.org/doc/html/rfc1630> (accessed: 25/11/2024).
- Böhm, Astrid, Julia Eibinger, Helmut W. Klug et al.:** CoReMA: Cooking Recipes of the Middle Ages. In: Roman Bleier and Helmut W. Klug (eds.): *Digitale Edition in Österreich [Digital/Scholarly Edition in Austria]*, Norderstedt 2023, pp. 159–160. urn:nbn:de:hbz:38-704790 (accessed: 25/11/2024).
- Burghart, Majorie (ed.):** Creating a Digital Scholarly Edition with the Text Encoding Initiative (2017). [https://ec.europa.eu/programmes/erasmus-plus/project-result-content/cfb03c8e-6765-477e-aa4d-e2b162f57068/IO2\\_Creating%20a%20digital%20edition%20with%20the%20TEI.pdf](https://ec.europa.eu/programmes/erasmus-plus/project-result-content/cfb03c8e-6765-477e-aa4d-e2b162f57068/IO2_Creating%20a%20digital%20edition%20with%20the%20TEI.pdf) (accessed: 19/12/2024).
- Burnard, Lou:** The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure. In: *Journal of the Text Encoding Initiative 5* (2013). <https://doi.org/10.4000/jtei.811> (accessed: 25/11/2024).
- Burrows, Toby, Antoine Brix, Douglas Emery et al.:** Linked Open Data Vocabularies and Identifiers for Medieval Studies. In: Sanita Reinsone, Inguna Skadiņa, Anda Baklāne et al. (eds.): *Digital Humanities in the Nordic Countries 2020. Post-Proceedings of the 5th Conference Digital Humanities in the Nordic Countries (DHN 2020)*. Riga, Latvia, October 21–23, 2020, pp. 211–218. <https://ceur-ws.org/Vol-2612/short5.pdf> (accessed: 25/11/2024).
- Burrows, Toby, Laura Cleaver, Doug Emery et al.:** Medieval Manuscripts and Their Migrations: Using SPARQL to Investigate the Research Potential of an Aggregated Knowledge Graph. In: *Digital Medievalist 15/1* (2022). <https://>

doi.org/10.16995/dm.8064  
(accessed: 25/11/2024).

- Burrows, Toby, Matthew Holford, David Lewis et al.:** Transforming TEI Manuscript Descriptions into RDF Graphs. In: Elena Spadini, Francesca Tomasi and Georg Vogeler (eds.): *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing* (Schriften des Instituts für Dokumentologie und Editorik 15). Norderstedt 2021, pp. 143–154.
- Dröge, Evelyn, Julia Iwanowa, and Steffen Hennicke:** A Specialization of the European Data Model for the Representation of Manuscripts: The DM2E Model. In: *Libraries in the Digital Age (LIDA) Proceedings*. 2014. [http://lida.ffos.hr/2014/files/LIDA\\_2014\\_34\\_Droegge.pdf](http://lida.ffos.hr/2014/files/LIDA_2014_34_Droegge.pdf) (accessed: 19/12/2024).
- Eide, Øyvind and Christian-Emil Smith Ore:** Ontologies and Data Modeling. In: Julia Flanders and Fotis Jannidis (eds.): *The Shape of Data in the Digital Humanities*. London 2018, pp. 178–203. <https://doi.org/10.4324/9781315552941> (accessed: 25/11/2024).
- Gehrke, Stefanie, Eduard Frunzeanu, Pauline Charbonnier et al.:** *Biblistima's* Prototype on Medieval Manuscript Illuminations and Their Context. In: Arnaud Zucker, Isabelle Draelants, Catherine Faron Zucker et al. (eds.): *Proceedings of the First International Workshop Semantic Web for Scientific Heritage at the 12th ESWC 2015 Conference*. 2015, pp. 43–48. <https://ceur-ws.org/Vol-1364/paper5.pdf> (accessed: 25/11/2024).
- Guarino, Nicola:** Formal Ontology and Information Systems. In: Idem (ed.): *Formal Ontology in Information Systems*. Proceedings of FOIS'98, Trento, Italy, 6–8 June 1998. Amsterdam 1998, pp. 3–15.

Handschriftenportal. 2018. <https://handschriftenportal.de/>  
(accessed: 25/11/2024).

- Heath, Tom and Christian Bizer:** Linked Data: Evolving the Web into a Global Data Space. In: *Synthesis Lectures on the Semantic Web. Theory and Technology 1* (2011), pp. 1–136. <http://dx.doi.org/10.2200/S00334ED1V01Y201102WBE001> (accessed: 25/11/2024).
- Jordanous, Anna, Mark Hedges, Faith Lawrence et al.:** Exploring Manuscripts: Sharing Ancient Wisdoms across the Semantic Web. In: *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS '12)*. New York 2012, pp. 1–12. <https://doi.org/10.1145/2254129.2254184> (accessed: 25/11/2024).
- Rehbein, Malte:** Ontologien. In: Fotis Jannidis and Hubertus Kohle: *Digital Humanities*. Stuttgart 2017, pp. 162–176.
- Sahle, Patrick, Georg Vogeler, Jana Klinger et al. (eds.):** *A Catalog of Digital Scholarly Editions, Version 4.107*. 2020– (last change 2024-04-17). <https://www.digitale-edition.de/> (accessed: 25/11/2024).
- Schaßan, Torsten:** The Use of TEI in the Handschriftenportal. In: *TEI2022 Conference and Members' Meeting "Text as Data"*, Newcastle, UK. Zenodo (2022, 19 September). <https://doi.org/10.5281/zenodo.7092430> (accessed: 25/11/2024).
- Steiner, Christian and Robert Klugseder:** Textmodellierung und Analyse von quasi-hierarchischen und varianten Liturgika des Mittelalters. In: Roman Bleier, Franz Fischer, Torsten Hiltmann et al. (eds.): *Digitale Mediävistik (Das Mittelalter. Perspektiven mediävistischer Forschung 24)*. Berlin 2019, pp. 205–220. <https://doi.org/10.1515/mial-2019-0014> (accessed: 25/11/2024).

- Stifter, David:** The Language of the Poems of Blathmac. In: Pádraig Ó Riain (ed.): The Poems of Blathmac Son of Cú Brettan. Reassessments (Irish Texts Society, Subsidiary Series 27). London 2015, pp. 47–103.
- Vieira, Jose Miguel and Arianna Ciula:** Implementing an RDF/OWL Ontology on Henry the III Fine Rolls. In: Christine Golbreich, Aditya Kalyanpur and Bijan Parsia (eds.): Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions. 2007. <http://ceur-ws.org/Vol-258/paper06.pdf> (accessed: 25/11/2024).
- Vogeler, Georg:** The ‘Assertive Edition’. In: International Journal of Digital Humanities 1 (2019), pp. 309–322. <https://doi.org/10.1007/s42803-019-00025-5> (accessed: 25/11/2024).
- Weidenmann, Arman:** Mittelalterliche Handschriften im Medienwandel. Historische Sammlungen zwischen Beharrungswillen und Innovationsdruck. In: Informationswissenschaft. Theorie, Methode und Praxis 6 (2020), pp. 91–149. <https://doi.org/10.18755/iw.2020.7> (accessed: 25/11/2024).
- Witt, Jeffrey C.:** Digital Scholarly Editions and API Consuming Applications. In: Roman Bleier, Martina Bürgermeister, Helmut W. Klug et al. (eds.): Digital Scholarly Editions as Interfaces (Schriften des Instituts für Dokumentologie und Editorik 12). Norderstedt 2018, pp. 219–247.
- Zhao, Fudie:** A Systematic Review of Wikidata in Digital Humanities Projects. In: Digital Scholarship in the Humanities 38 (2023), pp. 852–874. <https://doi.org/10.1093/llc/fqac083> (accessed: 25/11/2024).
- Zhitomirsky-Geffet, Maayan and Gila Prebor:** Toward an Ontopedia for Historical Hebrew Manuscripts. In: Frontiers in Digital Humanities 3 (2016). <https://doi.org/10.3389/fdigh.2016.00003> (accessed: 25/11/2024).

# Die digitale Edition als Schnittstelle

## Editorische Workflows im KI-Zeitalter

### Kontakt

**Dr. Michael Schonhardt**,  
Akademie der Wissenschaften  
und der Literatur, Mainz,  
Institut für Sprach- und  
Literaturwissenschaft,  
Residenzschloss 1,  
64283 Darmstadt,  
michael.schonhardt@tu-darmstadt.de  
 <https://orcid.org/0000-0002-2750-1900>

**Abstract** Medieval studies are facing the challenge of large-scale digitisation of sources, similar in scale to the monumental editorial efforts of the 19th century. Beyond simple scanning, true digital transformation requires converting images into machine-readable, multi-layered data to meet the growing demands of Artificial Intelligence (AI)-driven tools like Large Language Models (LLM) and Automated Text Recognition (ATR). This paper highlights the importance of digital editions in this process, which need to be conceptualised as dynamic data generators rather than static products, using the digital edition of the ‘Decretum Burchardi’ as an example. Its workflow integrates International Image Interoperability Framework (IIIF)-mediated image data, semi-automated layout recognition, and transcription with ATR models, ensuring data reusability for academic research and machine learning and securing relevance for future scholarship.

**Keywords** Decretum Burchardi; Deep Learning; Digital Medieval Studies; Edition

## 1 Digitalisierung als Jahrundertaufgabe

„I always wondered how he did it“<sup>1</sup>, beginnt Howard BLOCH seine Biographie des berühmten Abbé Jacques-Paul MIGNE und meint damit die 218 Bände

---

<sup>1</sup> Bloch 1994, S. 1.

umfassende Editionssammlung ‚Patrologia Latina‘ (PL) aus dem 19. Jahrhundert, neben der er noch acht weitere Serien mit insgesamt mehreren hundert Bänden druckte. Damit war er nicht allein, brachte das 19. Jahrhundert doch eine Vielzahl beeindruckender Editionsprojekte hervor, die noch heute das Rückgrat der historischen Forschung bilden. In Deutschland etwa die ‚Monumenta Germaniae Historica‘ (MGH), deren Herausgeber mit großem Einsatz die Veröffentlichung der Quellen des Mittelalters vorantrieben, wovon unzählige Handschriftenreisen in die entlegensten Ecken Europas zeugen, wie sie zum Beispiel im ‚Deutschen Archiv für Erforschung des Mittelalters‘ belegt sind.<sup>2</sup>

Natürlich ist die Sammlung und Verbreitung historischer Quellen kein Alleinstellungsmerkmal des 19. Jahrhunderts, sondern spätestens seit dem Humanismus Teil historisch-philologischer Gelehrsamkeit.<sup>3</sup> Gleichwohl war das Unterfangen, das in MIGNES *atelier catholique* und bei den MGH verfolgt wurde, durchaus revolutionär: Erst die Überführung der verstreuten Überlieferung in eine standardisierte Ausgabe, die in allen größeren Bibliotheken simultan zur Verfügung stand, erlaubte Zitierbarkeit, Auffindbarkeit und den Ansatz von Vollständigkeit des Quellenmaterials – eine unerlässliche Vorbedingung für die Etablierung der Mediävistik als moderne Wissenschaft.<sup>4</sup>

Dies war nicht weniger als eine Jahrhundertaufgabe, deren Erfolg vor allem zwei Grundlagen hatte: einerseits ein verändertes Selbstverständnis von Editorik, die sich zunehmend als methodengeleitete Dienstleistung verstand,<sup>5</sup> andererseits den technologischen Wandel, der erst im Laufe des 19. Jahrhunderts solche Vorhaben möglich machte. So griff MIGNÉ nicht nur auf fast 600 Angestellte zurück, sondern auch auf fünf dampfbetriebene Druckmaschinen, die rund um die Uhr produzierten.<sup>6</sup> Dazu erleichterten neue Reise-, Versand- und Reproduktionsmöglichkeiten Akquise und Verbreitung von bislang unzugänglichen Quellen und Forschungsergebnissen.<sup>7</sup>

Rund zwei Jahrhunderte nach diesen Erfolgen steht die Mediävistik erneut einer Jahrhundertaufgabe gegenüber, die deutliche Parallelen zu den historischen Vorhaben aufweist: der umfassenden Digitalisierung der Quellen im Windschatten

---

2 Zur Entstehungsgeschichte der MGH vgl. Bünz 2019. Zu den frühen Bedingungen dieser Vorhaben und ihrer Parallelität zu heute auch Schonhardt 2014.

3 Dies bezeugt die enge Bindung vieler Humanisten an den Druckhandel; vgl. etwa Davies 1996; Holeczek 1993; Muhlack 1993.

4 Die Auswirkungen dieser neuen Wissensfülle beschreibt Blair 2010.

5 Zur zentralen Rolle Lachmanns, dessen Methode heute zu gering geschätzt wird, vgl. Weigel 1989; Trovato u. Reeve 2014. Zum weiteren methodologischen Umfeld vgl. Baker 2011, hier S. 429–431.

6 Bloch 1994, S. 14f.

7 Zum Wandel der Reproduktionstechnik vgl. McKitterick 2014, S. 72–138.

der gesamtgesellschaftlichen Digitalisierung.<sup>8</sup> Nahmen sich dieser Herausforderung in den letzten zwei Dekaden vor allem Bibliotheken und Archive durch das Scannen von Akten und Handschriften an,<sup>9</sup> so wird heute deutlich, dass Bilddigitalisierung nur ein erster Schritt sein kann. Denn obwohl eine Quelle dadurch technisch digitalisiert ist, ist ihre Rezeption am Bildschirm weiterhin auf die analoge Kulturtechnik des linearen Lesens beschränkt und damit mit Blick auf ihren Quellengehalt dem Menschen vorbehalten.<sup>10</sup> Die Maschine kann lediglich Pixel-Informationen verarbeiten, der im fotografierten Text codierte Inhalt bleibt ihr verschlossen und wird erst durch Textdigitalisierung im kulturtechnischen Sinne digital.<sup>11</sup>

Das volle Potential von Digitalisierung erfüllt sich daher erst durch die Überführung der erzeugten Bilder in maschinenlesbare Datenrepräsentationen von Text, Inhalt und weiteren Aspekten, welche zum Beispiel die kodikologischen oder paläographischen Eigenschaften der Quellen betreffen. Aktuelle Entwicklungen im Bereich der Künstlichen Intelligenz (KI), vor allem des *Deep Learnings* (DL), steigern die Notwendigkeit solcher Daten rasant: Mittlerweile sind Maschinen zu menschenähnlichen Leistungen im Bereich der Informationsextraktion, beim Entziffern von Handschriften oder im Umgang mit natürlicher Sprache fähig und können somit auch in der geisteswissenschaftlichen Forschung eingesetzt werden.<sup>12</sup>

Gleichzeitig erhöht diese Technologie die Notwendigkeit einer Digitalisierung, die der Maschine einen Zugang zu wissenschaftlichen Texten und Quellen ermöglicht. Einerseits, da Sprachmodelle und andere Anwendungen des DL, etwa die mittlerweile etablierte *Automated Text Recognition* (ATR), auf enorme Mengen an Daten angewiesen sind, um bestimmte Fähigkeiten adäquat zu erlernen;<sup>13</sup> andererseits, da sich diese Fähigkeiten natürlich nur auf Daten anwenden lassen, die im Rahmen des Trainings oder durch *Retrieval-Augmented Generation* (RAG) zugänglich sind.<sup>14</sup>

Es ist abzusehen, dass KI auch in den Wissenschaften einen festen Platz einnehmen wird, so wie heute „Lesen und Schreiben am Bildschirm, Kommunikation über Emails, Nutzung digitaler Wörterbücher, digitalisierter Quellen etc. den Alltag nahezu aller geisteswissenschaftlichen Arbeit bestimmt“<sup>15</sup>. Möchte

---

8 Vgl. Terras 2018.

9 Vgl. Klaffki, Schmunk u. Stäcker 2018.

10 Vgl. Sahle 2013b, S. 58 f. Zur Digitalisierung als Kulturtechnik vgl. Krämer 2018; Krämer 2021.

11 Hierzu Rehbein 2017, hier S. 193–196; Jannidis 2017.

12 Vgl. Naveed, Khan, Qiu u. a. 2023. Zur Anwendung im geschichtswissenschaftlichen Kontext vgl. Hodel 2024.

13 Zum Datenbedarf von ATR vgl. Hodel, Schoch, Schneider u. a. 2021, hier S. 7. Zur Größe von Trainingssets für LLMs vgl. Brown, Mann, Ryder u. a. 2020, S. 8 f.

14 Zu RAG vgl. Gao, Xiong, Gao u. a. 2023.

15 Krämer 2018, S. 6.

auch die Mediävistik an diesen Entwicklungen partizipieren, muss sie ihre Quellen- und Forschungsdaten in angemessener Weise zur Verfügung stellen, damit sie zum Beispiel durch digitale Assistenten oder andere Anwendungen rezipiert werden können.

Eine solche aktive Partizipation wäre auch deswegen zu wünschen, da durch DL-Technologien wie die *Large Language Models* (LLMs)<sup>16</sup> eben nicht die oft befürchtete „evidenzorientierte Kolonialisierung der Geisteswissenschaften durch positivistische Erkenntnismethoden“<sup>17</sup> droht. Die Ablösung algorithmischer Programmierung als einziges Kommunikationsmedium mit der Maschine durch „Interpretationskompetenz“<sup>18</sup> eröffnet vielmehr gerade den Geisteswissenschaften nachhaltige Relevanzperspektiven,<sup>19</sup> sofern sich diese durch eine auf Vollständigkeit zielende Datafizierung ihrer Gegenstände daran beteiligen.

## 2 Die (digitale) Edition als Schnittstelle

Entsprechend steht die Mediävistik daher erneut vor einer Jahrhundertaufgabe, die den Vorhaben des 19. Jahrhunderts in nichts nachsteht und in der Überführung des Quellenbestands in eine den technischen und praxeologischen Anforderungen ihrer Zeit entsprechenden Medienform liegt. Damals wie heute muss diese enorme Aufgabe von Editionsprojekten geleistet werden:<sup>20</sup> Erstens, weil die penible Aufbereitung von Daten seit jeher zur Kernaufgabe einer Edition gehört und keinen wesentlichen Mehraufwand bedeutet.<sup>21</sup> Zweitens, weil die Editorik bereits seit Jahren Erfahrung in der digitalen Erschließung ihres Materials gesammelt hat.<sup>22</sup> Und drittens, weil editorische Verfahren nicht nur auf die Erstellung deskriptiver Transkriptions- und Kollationsdaten zielen, sondern auch die sachkundige Prüfung verschiedenster Phänomene einschließen.<sup>23</sup>

Damit Editionen aber dieser Schlüsselfunktion für die digitale Transformation der Mediävistik gerecht werden können, müssen sie die Datafizierung ins Zentrum ihres Selbstverständnisses stellen.

---

16 Zu aktuellen GPTs (*Generative Pre-Trained Transformers*) vgl. Bubeck, Chandrasekaran, Eldan u. a. 2023; Wolfram 2023.

17 Krämer 2018, S. 7.

18 Ebd.

19 Zum Prompt-Engineering vgl. Wei, Wang, Schuurmans u. a. 2022.

20 Zur zentralen Rolle von Bibliotheken aber Schaßan 2013.

21 Vgl. hierzu West 1973; Delz 1997.

22 Vgl. Earhart 2012, S. 18–28.

23 Vgl. Schaßan 2013, S. 299; Sahle 2013a, S. 55.

[Dies] bedeutet für alle nicht ‚digital geborenen‘ Materialien, dass diese in einer Weise codiert werden müssen, die grundsätzliche Entscheidungen seitens der Forschenden verlangt, was jeweils zu markieren, zu annotieren ist, wie Informationsgehalte zu hierarchisieren sind, in dem Sinne etwa, was als ‚Text‘ und was als ‚Metadaten‘ aufzufassen ist und ähnliches mehr. ‚Verdatung‘ heißt also nicht einfach etwas Vorliegendes einzuscannen; vielmehr verwandelt sich das textuelle Ausgangsobjekt, insofern im Objekt angelegte Implikationen bei seiner Verdatung explizit gemacht werden müssen.<sup>24</sup>

So verstanden, ist Datafizierung seit jeher Kernaufgabe der Editorik, die auch im Druck den handschriftlichen Befund der Überlieferung durch editorische Hilfsmittel wie Apparate, Register, Symbole und erläuternde Fußnoten in formalisierter und expliziter Weise nach Maßgabe eines Datenmodells transparent und strukturiert aufbereitet,<sup>25</sup> allerdings entsprechend der zweidimensionalen und statischen Natur der gedruckten Seite als flaches Datenmodell.<sup>26</sup>

Zentrales Manko analoger Editionen ist daher gerade ihre Langlebigkeit, die den einmal gedruckten Text auf viele Jahr(hundert)e zugänglich hält, dabei aber mit seiner Endgültigkeit einhergeht. Somit sind Editionen, die als stabiles Produkt konzipiert sind, nicht in der Lage, den veränderten Anforderungen digitaler Nutzungspraktiken zu entsprechen, selbst wenn sie als Retrokonversion digital vermittelt vorliegen.<sup>27</sup>

Weit größeres Potential bieten hier digitale Editionen, die eine ‚digitale‘ Datafizierung ermöglichen, ihren Gegenstand also in elektronische Systeme mit einem mehrdimensionalen Informationsraum überführen.<sup>28</sup> Bild-, Text-, Struktur-, Meta- und andere codierte Markupdaten lassen sich so als geschichtete Datenstrukturen aufbewahren, die erst bei der Ausgabe am Bildschirm dynamisch visualisiert oder in größere Zusammenhänge aggregiert werden.<sup>29</sup>

Trotz dieser potentiellen Dynamik sind aber auch digitale Editionen häufig als geschlossenes und auf Stabilität zielendes Produkt nach der Maßgabe antizipierter Nutzungsszenarien und Forschungsinteressen angelegt.<sup>30</sup> Während die Nutzung durch menschliche Akteure aber aufgrund der in einem Fachgebiet etablierten

<sup>24</sup> Krämer 2018, S. 6.

<sup>25</sup> Vgl. z. B. West 1973, S. 72–102.

<sup>26</sup> Vgl. Sahle 2013a, S. 270–289.

<sup>27</sup> Vgl. Shillingsburg 1996.

<sup>28</sup> Vgl. Krämer 2018, S. 9.

<sup>29</sup> Vgl. Pierazzo 2015, S. 25–29.

<sup>30</sup> Vgl. ebd., S. 2 f.; Sahle 2013a, S. 270, beschreibt dies als Printparadigma, das auch im digitalen Raum vorherrschen kann.

und bekannten Praktiken verhältnismäßig stabil bleibt, führt der oben skizzierte technische Wandel zu wechselnden maschinellen Akteuren, deren Anforderungen an die Daten einer Edition von menschlichen Bedürfnissen abweichen.<sup>31</sup>

Ein anschauliches Beispiel hierfür ist die ATR, die sich in den verbreiteten Tools eScriptorium und Transkribus bereits großer Beliebtheit erfreut. Hier wird durch den massenhaften Abgleich von Bilddateien und dazugehörigen Transkriptionen eine Repräsentation – genannt Modell – der erwünschten Transkriptionsweise im Verhältnis zum Schriftbefund generiert, die dann auf neues Bildmaterial appliziert werden kann.<sup>32</sup> Eine Herausforderung besteht in der Ansammlung ausreichender Mengen an Trainingsmaterial, das teuer und aufwendig manuell transkribiert werden muss. Editionen wären hier ein idealer Datenlieferant ohne nennenswerten Mehraufwand.<sup>33</sup> Hierzu werden diese Transkriptionen aber in einem spezifischen Datenformat benötigt (in der Regel PageXML<sup>34</sup>), das vor dem rezenten Durchbruch der ATR im Datenportfolio digitaler Editionen eine untergeordnete Rolle gespielt hat. Zwar lassen sich solche Datenformate technisch aus bestehenden Strukturen wie der *Text Encoding Initiative* (TEI)<sup>35</sup> transformieren, aber nur, wenn dort Bild-, Layout- und Textdaten aufgenommen und miteinander referenziert wurden. Auch dann ist ein Nutzungspotential digitaler Editionsdaten im Rahmen der ATR aber nicht garantiert. Neben den Anforderungen, die im Hinblick auf das Format zu erfüllen sind, müssen auch die Inhalte hierfür geeignet sein:<sup>36</sup> Da ATR ein stochastisches Abbild der zugeführten Daten erzeugt, ist ihre Transkriptionsfähigkeit auf die in den Daten umgesetzte Transkriptionspraxis beschränkt. Einmal ins Modell trainierte Zeichensätze lassen sich daher nicht ohne weiteres in andere überführen.<sup>37</sup>

Daher können aus Editionen generierte Daten nur für das Training von ATR-Modellen genutzt werden, die den gleichen Richtlinien entsprechen. Aufgrund der unterschiedlichen Praktiken innerhalb verschiedener Projekte und Disziplinen ist es aber nahezu unmöglich, diese im Rahmen einer editorischen Aufarbeitung

---

31 Vgl. hierzu auch Ralle 2016.

32 Das Modell ‚CATMuS Medieval‘ wurde etwa an 113.228 Textzeilen trainiert; vgl. Clérice, Pinche, Vlachou-Efstathiou u. a. 2024.

33 Vgl. zu den Grundlagen der ATR Hodel 2022, S. 65–80; Pinche 2023; Clérice, Pinche, Vlachou-Efstathiou u. a. 2024.

34 Siehe <https://www.digitale-edition.at/o/konde.154> (Zugriff: 08.06.2024).

35 Siehe <https://www.digitale-edition.at/o/konde.178> (Zugriff: 08.06.2024).

36 Zur Nachnutzbarkeit vgl. Romein, Hodel, Gordijn u. a. 2022; Guéville u. Wrisley 2022. Eine wichtige Ressource für die Nachnutzung von ATR-Daten ist ‚HTR United‘: <https://htr-united.github.io> (Zugriff: 08.06.2024). In jüngster Zeit wird die Standardbildung vor allem in der französischen Forschung vorangetrieben; vgl. Clérice, Vlachou-Efstathiou u. Chagué 2023, S. 4.

37 Vgl. Hodel 2023, hier S. 157 f.

adäquat zu antizipieren. Die Nachnutzbarkeit spielt daher nur selten eine Rolle bei den Transkriptionsentscheidungen digitaler Editionen, die eher mit Blick auf die Vorlieben menschlicher Rezipienten und knappe Ressourcen getroffen werden.<sup>38</sup> In der Folge mangelt es auch heute noch an Daten für das Training generischer Modelle von hoher Qualität.<sup>39</sup>

Dabei ist ATR nur ein Nachnutzungsfeld editorischer Daten unter vielen. Da DL-Modelle prinzipiell jedes Muster abbilden können – etwa kodikologische, paläographische, philologische oder inhaltliche Phänomene –, ist es unmöglich, zukünftige Anforderungen umfassend vorherzusehen. Vielmehr müssen Editionen einen Weg finden, die zunehmend situativen, dynamischen und technikgebundenen Anforderungen an das eigene Datenprofil zu erfüllen, ohne sie im Einzelnen antizipieren zu können. Hierzu ist es notwendig, die Edition konzeptionell vom statischen Produkt (sei es im Druck, als PDF oder Website) zum dynamischen Prozess weiterzudenken. Dabei bezieht sich diese Prozesshaftigkeit nicht nur auf den Status des edierten Texts im Sinne eines *editorial laboratory*,<sup>40</sup> sondern auch auf eine technische Prozesshaftigkeit als Schnittstelle und Datengenerator, in der sich editorische Konzepte, Workflows und deren technische Umsetzung gegenseitig bedingen. Dieses Konzept soll im Folgenden anhand der Edition des ‚Decretum Burchardi‘ im Rahmen des Langzeitprojekts ‚Burchards Dekret Digital‘ gezeigt werden.

### 3 Von der Handschrift zur Schnittstelle: Das Praxisbeispiel ‚Burchards Dekret Digital‘

Im Zentrum dieses Projekts steht mit dem ‚Decretum Burchardi‘ eine kirchenrechtliche Sammlung des 11. Jahrhunderts. Dessen Urheber, Burchard von Worms, war an der Wende vom 10. zum 11. Jahrhundert Bischof von Worms und kompilierte in diesem Zusammenhang eine Sammlung verschiedener Rechtssätze des kirchlichen Rechts, die sich jeweils einem bestimmten Thema widmen.<sup>41</sup> Im Vergleich zur allgemeinen Überlieferung mittelalterlicher Texte zeichnet sich Burchards Dekret durch eine fast singuläre Situation aus. So konnten eine Reihe von Handschriften identifiziert werden, die noch zu Lebzeiten des Bischofs in Worms entstanden sind und in ihrem kodikologischen Befund deutliche Spuren eines komplexen und mehrschichtigen Entstehungs- und Entwicklungsprozesses

---

38 Zu den verschiedenen Transkriptionspraktiken vgl. Vander Meulen u. Tanselle 1999; Driscoll 2007; Robinson u. Solopova 1993; Guéville u. Wrisley 2022.

39 Hodel, Schoch, Schneider u. a. 2021.

40 Vgl. Dahlström 2004.

41 Vgl. Austin 2009; Austin 2019, S. 458–470.

aufweisen.<sup>42</sup> Allerhand Rasuren, Umstellungen, Durchstreichungen, Ergänzungen und andere Zeugnisse stattgefundener Überarbeitungen bieten dort einen seltenen Blick auf den Schreibtisch eines mittelalterlichen Kanonisten. Gleichzeitig stellt das Dekret und seine Überlieferung aber auch eine reichhaltige Fundgrube dar, die über den kanonistischen Fokus hinausreicht und in Zukunft für verschiedene wissenschaftliche Kontexte Relevanz entfalten kann. Burchards Dekret ist somit ein vielschichtiges Gebilde, das im Rahmen der digitalen Edition sorgsam zu erschließen ist.

Auch der enorme Umfang des Dekrets, das in den wichtigsten Handschriften ca. 150.000 Zeilen umfasst, in denen sich die verschiedensten Erscheinungen in großer Zahl finden lassen, steigert sein Nachnutzungspotential.<sup>43</sup> Das Dekret liefert somit eine wertvolle Datengrundlage für Training und Anwendung der oben skizzierten Digitalverfahren von der ATR über LLMs bis hin zu spezialisierteren KI-Anwendungen für die Erkennung von Rasuren, Schreiberwechseln oder anderen Phänomenen. Das Potential der erhobenen Daten ist damit enorm, führt aber auch zu einem entsprechenden Anspruch an die digitale Edition und ihre Workflows. Sie müssen nicht nur Dinge erschließen, die heute als relevant erscheinen, sondern auch konzeptionell und technisch unvorhersehbare Anwendungsszenarien im Rahmen des Machbaren ermöglichen. Hierbei helfen vier Axiome, die der Entwicklung des editorischen Workflows zugrunde liegen:

- 1) Zukünftige Anforderungsprofile werden sich von aktuellen unterscheiden.
- 2) Zukünftige Verfahren werden ihren Ausgang von Text-, Bild- oder jeweils davon abgeleiteten Annotationsdaten nehmen.
- 3) Grundeinheiten dieser Verfahren sind die Handschriftenseite und Textzeilen.
- 4) Zukünftige Verfahren werden diese Datenstrukturen statisch (zum Beispiel durch einmalige Erstellung einer neuen oder alternativen Webansicht) oder situativ (zum Beispiel durch multimodale technische Agenten<sup>44</sup>) aufeinander beziehen und durch externe Daten erweitern (zum Beispiel durch im Nachgang erhobene Daten einer Multispektralanalyse<sup>45</sup>).

Diese Axiome stellen den konzeptionellen Ausgangspunkt eines Workflows dar, der in erster Linie die Verbindung von Text- und Bilddaten sowie deren

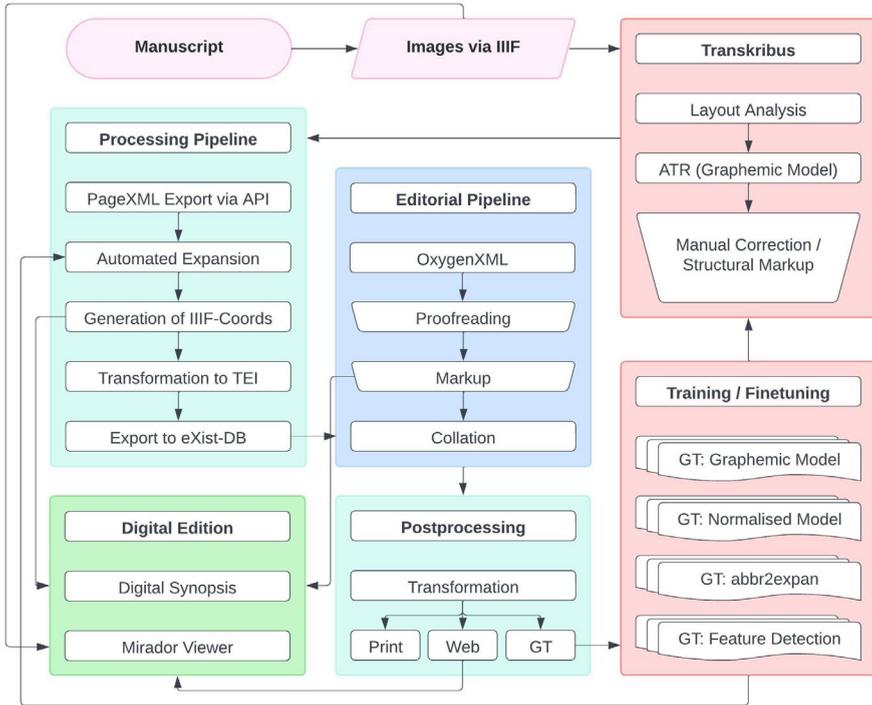
---

<sup>42</sup> Vgl. hierzu Hoffmann u. Pokorny 1991.

<sup>43</sup> Bamberg, Staatsbibliothek, Msc. Can. 6 (312 Blatt); Frankfurt, Universitätsbibliothek, Ms. Barth. 50 (315 Blatt); Köln, Erzbischöfliche Diözesan- und Dombibliothek, Cod. 119 (204 Blatt); Vatikan, Biblioteca Apostolica Vaticana, Pal. lat. 585 und 586 (330 und 280 Blatt).

<sup>44</sup> Vgl. Woolridge u. Jennings 1995. Zu multimodalen Systemen vgl. Barua, Ahmed u. Begum 2023.

<sup>45</sup> Vgl. etwa Shevchuk, Bonnerot, Brockmann u. a. 2021.



**Abb. 1** | Workflow des Projekts ‚Burchards Dekret Digital‘. Bild des Autors, CC BY 4.0.

Nachnutzbarkeit im Sinne der FAIR-Kriterien (*Findable, Accessible, Interoperable, Reusable*) garantieren soll (Abb. 1).<sup>46</sup>

Er nimmt seinen Ausgang in den *International Image Interoperability Framework* (IIIF)<sup>47</sup>-vermittelten Bilddaten der Textzeugen, die in Transkribus einer automatischen Layouterkennung und -analyse unterzogen werden.<sup>48</sup> Der so erkannte Textbereich wird dann durch ein graphematisches Modell erkannt, das an ca. 100.000 transkribierten Wörtern trainiert wurde (*Character Error Rate* [CER] 2,4 Prozent).<sup>49</sup> Nach manueller Korrektur erfolgt die automatisierte

<sup>46</sup> Vgl. Wilkinson, Dumontier, Aalbersberg u. a. 2016.

<sup>47</sup> Siehe <https://www.digitale-edition.at/o:konde.123> (Zugriff: 08.06.2024).

<sup>48</sup> Diese Layoutdaten sind in Schonhardt 2024b veröffentlicht und als Kraken-Modelle (Schonhardt 2024c sowie Schonhardt 2024a) nachnutzbar.

<sup>49</sup> Dieses Modell steht unter <https://www.transkribus.org/model/bdd-wormser-scriptorium-0.2> zur Nachnutzung in Transkribus bereit, ebenso als auflösendes Modell <https://www.transkribus.org/model/bdd-wormser-scriptorium-expanded-0.1>. Beide Modelle wurden in Kraken nachtrainiert (<https://zenodo.org/records/13741957> und <https://zenodo.org/records/13736584>) (Zugriff: 08.06.2024).

Überführung von Layout- und Textdaten aus PageXML in projektkonformes TEI. Dies beinhaltet eine automatische Auflösung der Abkürzungen, IIF-konforme Integration der Layout-Koordinaten in TEI, eine automatische Strukturauszeichnung sowie den Export in eine interne Arbeitsdatenbank (eXist-db).<sup>50</sup>

Im XML-Editor OxygenXML werden weiterführende Auszeichnungen vorgenommen, die erzeugten Daten manuell geprüft und anschließend in verschiedene Datenstrukturen transformiert. Diese ermöglichen zum einen die Visualisierung der Edition im Web<sup>51</sup> und im Druck, können zum anderen aber auch als Trainingsdaten für weitere DL-Verfahren aufbereitet bzw. in entsprechende Modelle überführt werden. Dies umfasst aktuell verschiedene Transkriptionsdaten (graphematisch und normalisiert) sowie ein Datenset zum Training zur automatisierten Auflösung von Abkürzungen in einer Transformer-Architektur.<sup>52</sup> Möglich ist aber auch die Erzeugung von *Ground Truth* für andere Anwendungsbereiche, etwa für die Erkennung von bestimmten kodikologischen Phänomenen wie Rasuren oder Schreiberwechsel, da alle ausgezeichneten Phänomene einer Bildkoordinate zugeordnet werden können.

Der hier skizzierte Workflow erzeugt daher nicht nur eine TEI-codierte Wiedergabe der Textgestalt des Dekrets, sondern mündet entsprechend der oben genannten Axiome in eine ganze Reihe von Datenstrukturen:<sup>53</sup>

- 1) IIF-Manifeste
- 2) Bilddaten
- 3) Layout- und Transkriptionsdaten
- 4) Annotationsdaten
- 5) TEI-Codierungen von Transkription, Edition, Register und Listen
- 6) *Ground Truth* und Modelle für DL-Verfahren
- 7) Präsentationsdaten (HTML, CSS, JS)

Dabei wird ein integrativer Ansatz verfolgt, in dem alle Datenstrukturen durch automatisiert vergebene *Identifier* der übergreifenden Elemente wie Handschriftenseiten, Spalten, Zeilen und Marginalien verbunden sind. Hierdurch können unterschiedliche Datenstrukturen sinnvoll aufeinander bezogen und in neue

---

<sup>50</sup> Skripte unter <https://github.com/michaelscho/transpy> (Zugriff: 08.06.2024).

<sup>51</sup> Unter <https://www.burchards-dekret-digital.de> (Zugriff: 08.06.2024).

<sup>52</sup> Skripte <https://github.com/michaelscho/MediAbbr> (Zugriff: 08.06.2024). Der Ansatz basiert auf der im Bereich der maschinellen Übersetzung verbreiteten ‚T5 Architektur‘. Vgl. Raffel, Shazeer, Roberts u. a. 2019. Zu Transformern vgl. Vaswani, Shazeer, Parmar u. a. 2017.

<sup>53</sup> Unter <https://gitlab.rlp.net/adwmainz/projekte/burchards-dekret-digital/data> (Zugriff: 08.06.2024) sowie <https://gitlab.rlp.net/adwmainz/projekte/burchards-dekret-digital/website> (Zugriff: 08.06.2024).

Zusammenhänge transformiert werden. Beispielsweise nutzt das Projekt diese Verbindungen, um aus den manuell korrigierten und in TEI codierten Transkriptionen automatisiert Daten für das Training verschiedener ATR- und Normalisierungsmodelle zu gewinnen. Prinzipiell lassen sich auf diese Weise alle editorisch aufbereiteten Phänomene auf dem Bild der Handschrift verorten und somit in KI-gestützte Analysekontexte überführen.<sup>54</sup> Vermittelt werden diese Datenschichten durch IIF-Manifeste, die neben der rechtskonformen Einbindung von Digitalisaten auch die generierten Datenrepräsentationen verlinken, die zur jeweiligen Handschriftenseite gehören.

Die Präsentation der Edition erfolgt dann als gedruckte Ausgabe und als Webedition.<sup>55</sup> Diese wird als statisches HTML generiert, dessen dynamische Prozessierung durch Javascript im Browser erfolgt und ohne komplexe technische Schnittstelle auskommt. Das Herz dieser Edition stellt der sogenannte Schreibtisch dar, auf dem die erzeugten Daten nach Maßgabe des eigenen Forschungsinteresses individuell zusammengestellt werden können.<sup>56</sup> Hier können nicht nur Transkriptionen abgerufen werden, sondern auch deren Strukturdaten sowie die Ansicht der jeweiligen Handschrift. Außerdem können die Textzeugen einer Synopse unterzogen werden. Die Edition ist somit aus der Perspektive der Nutzenden als Prozess konzipiert, um eine wissenschaftliche Nutzung im Sinne eines *editorial laboratory* zu ermöglichen.<sup>57</sup> Die dabei entstehenden situativen Daten-Konfigurationen lassen sich dauerhaft als *JavaScript Object Notation* (JSON)-Dateien speichern und mit der Forschungscommunity über Repositorien oder in direkter Kommunikation teilen.

Damit weicht dieses Setup von der etablierten eXist-db-Infrastruktur ab,<sup>58</sup> die administrierte Webserver und Datenbanken voraussetzt, deren dauerhafte Pflege auch im Rahmen eines Langfristprojekts nur schwer zu garantieren ist. Daher greift das Projekt auf statisches HTML zurück, das über einen einfachen HTTP-Server – etwa GitLab Pages – gehostet und auf Zenodo<sup>59</sup> sowie Text+ (Nationale Forschungsdateninfrastruktur)<sup>60</sup> langzeitarchiviert werden kann, ohne Funktionalitäten zu verlieren. Datenarchivierung und Datenaustausch mit externen Zusammenhängen erfolgen über ein projekteigenes GitLab-Repositorium, das

---

54 Einen Anwendungsfall von Daten aus dem Bereich der Photographie, die aus IIF-konformen Annotationsdateien gewonnen wurden, liefern Cornut, Raemy u. Spiess 2023.

55 <https://www.burchards-dekret-digital.de> (Zugriff: 08.06.2024).

56 <https://burchards-dekret-digital.de/html/desktop.html> (Zugriff: 08.06.2024).

57 Dahlström 2004.

58 Auf eXist fußen z. B. die Editionsumgebungen Ediarum oder TEI-Publisher, vgl. Mertgens 2019.

59 <https://zenodo.org/> (Zugriff 23.10.2024).

60 <https://text-plus.org/> (Zugriff 23.10.2024).

im Projektverlauf auch auf Zenodo gesichert wird.<sup>61</sup> Auf diese Weise ermöglichen Workflow, Datenschema und Infrastruktur der Edition des Dekrets den vielseitigen Austausch und die fortwährende Rekombination erzeugter Daten.

#### 4 Fazit

Editionen bilden seit über 200 Jahren das Rückgrat mediävistischer Forschung. Hunderte von Ausgaben ermöglichen bis heute eine Zugänglichkeit, Vergleichbarkeit, Zitierbarkeit und Dauerhaftigkeit des Quellenbestands, der die Etablierung der Mediävistik als professionelle Wissenschaft erst ermöglichte. Daher kann es nicht verwundern, dass die Druckedition noch heute als Garant der Beständigkeit wissenschaftlicher Erkenntnis gilt, während der digitalen Edition oft der vermeintliche Makel prekärer Unsicherheit anhaftet.<sup>62</sup> Gleichzeitig ist die Rezeption dieser Quellen bislang menschlichen Akteuren vorbehalten, obgleich maschinelle Agenten zunehmend Funktionen im Wissensbildungsprozess einnehmen, die mit den wachsenden Fähigkeiten Künstlicher Intelligenz und dem damit im Zusammenhang stehenden Wandel der Kultur- und Wissenspraktiken weiter Relevanz entfalten werden. Digitale Editionen müssen damit bereits heute mehr leisten als die bloße Überführung der Quellen auf den Bildschirm. Vielmehr müssen sie im dauerhaften Austausch zwischen Überlieferung, Daten, menschlichen Akteuren und technischen Agenten die Datafizierung der Überlieferung des Mittelalters ermöglichen. Dabei besteht die Antwort auf die Frage: ‚How can they do it?‘ erneut in der aktiven Adaption moderner Technologien und der Aneignung eines neuen Verständnisses der Edition als generative und prozessuale Schnittstelle, damit aktuelle Editionen in 200 Jahren nicht nur schwer in den Regalen liegen, sondern sich – wie die Ausgaben von PL oder MGH – einer aktiven Nutzung durch Mensch und Maschine erfreuen.

---

<sup>61</sup> Siehe Anm. 53.

<sup>62</sup> Vgl. Pierazzo 2015, S. 4. Auch die von Steding 2001 festgehaltenen Vorbehalte gegenüber dem Digitalen sind noch heute gültig. Einen aktuellen Beitrag zum Verhältnis Druck- und Digitaledition bietet Kocher 2019.

## Literaturverzeichnis

- Austin, Greta:** Shaping Church Law around the Year 1000. The Decretum of Burchard of Worms (Church, Faith and Culture in the Medieval West). Farnham 2009.
- Austin, Greta:** Burchard of Worms. In: Philip L. Reynolds (Hg.): Great Christian Jurists and Legal Collections in the First Millennium. Cambridge 2019, S. 458–470.
- Baker, Craig:** Editing Medieval Texts. In: Albrecht Classen (Hg.): Handbook of Medieval Studies. Terms, Methods, Trends. Berlin 2011, S. 427–449.
- Barua, Arnab, Mobyen Uddin Ahmed u. Shahina Begum:** A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions. In: IEEE Access 11 (2023), S. 14804–14831.
- Blair, Ann:** Too Much to Know. Managing Scholarly Information Before the Modern Age. New Haven, London 2010.
- Bloch, R. Howard:** God's Plagiarist. Being an Account of the Fabulous Industry and Irregular Commerce of the Abbé Migne. Chicago 1994.
- Brown, Tom B., Benjamin Mann, Nick Ryder u. a.:** Language Models are Few-Shot Learners. 2020. <https://doi.org/10.48550/arXiv.2005.14165> (Zugriff: 08.06.2024).
- Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan u. a.:** Sparks of Artificial General Intelligence: Early Experiments with GPT-4. 2023. <https://doi.org/10.48550/arXiv.2303.12712> (Zugriff: 08.06.2024).
- Bünz, Enno:** Die Monumenta Germaniae Historica 1819–2019: Ein historischer Abriss. In: Mittelalter lesbar machen. Festschrift 200 Jahre Monumenta Germaniae Historica. Grundlagen, Forschung, Mittelalter. Wiesbaden 2019, S. 15–36.
- Clérico, Thibault, Ariane Pinche, Malamentenia Vlachou-Efstathiou u. a.:** CATMuS Medieval: A Multilingual Large-Scale Cross-Century Dataset in Latin Script for Handwritten Text Recognition and Beyond. 2024. <https://inria.hal.science/hal-04453952> (Zugriff: 08.06.2024).
- Clérico, Thibault, Malamentenia Vlachou-Efstathiou u. Alix Chagué:** CREMMA Medii Aevi: Literary Manuscript Text Recognition in Latin. In: Journal of Open Humanities Data 9 (2023). <https://doi.org/10.5334/johd.97> (Zugriff: 08.06.2024).
- Cornut, Murielle, Julien Antoine Raemy u. Florian Spiess:** Annotations as Knowledge Practices in Image Archives: Application of Linked Open Usable Data and Machine Learning. In: Journal on Computing and Cultural Heritage 16 (2023), S. 1–19.
- Dahlström, Mats:** How Reproductive is a Scholarly Edition? In: Literary and Linguistic Computing 19 (2004), S. 17–33.
- Davies, Martin:** Humanism in Script and Print in the Fifteenth Century. In: Jill Kraye (Hg.): The Cambridge Companion to Renaissance Humanism. Cambridge 1996, S. 47–62.
- Delz, Josef:** Textkritik und Editionstechnik. In: Fritz Graf (Hg.): Einleitung in die lateinische Philologie. Wiesbaden 1997, S. 51–73.
- Driscoll, Matthew James:** Electronic Textual Editing: Levels of Transcription. 2007. <https://tei-c.org/Vault/ETE/Preview/driscoll.html> (Zugriff: 08.06.2024).
- Earhart, Amy E.:** The Digital Edition and the Digital Humanities. In: Textual Cultures 7 (2012), S. 18–28.

**Gao, Yunfan, Yun Xiong, Xinyu Gao**

**u. a.:** Retrieval-Augmented Generation for Large Language Models: A Survey. 2023. <https://doi.org/10.48550/arXiv.2312.10997> (Zugriff: 08.06.2024).

**Guéville, Estelle u. David Joseph Wrisley:**

Transcribing Medieval Manuscripts for Machine Learning. 2022. <https://doi.org/10.48550/arXiv.2207.07726> (Zugriff: 08.06.2024).

**Hodel, Tobias:** Die Maschine und die Geschichtswissenschaft. In: Karoline Döring, Stefan Haas, Mareike König u. a. (Hgg.): *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*. Berlin 2022, S. 65–80.

**Hodel, Tobias:** Konsequenzen der Handschriftenerkennung und des maschinellen Lernens für die Geschichtswissenschaft. Anwendung, Einordnung und Methodenkritik. In: *Historische Zeitschrift* 316 (2023), S. 151–180.

**Hodel, Tobias:** Large Language Models, oder weshalb wir künstliche Intelligenz im Archiv finden sollten. In: Daniel Fährle u. Peter Müller (Hgg.): *Smart und intelligent – Digitale Unterstützung für die Arbeit im Archiv*. Ostfildern 2024, S. 77–84. <https://boris.unibe.ch/197467/> (Zugriff: 08.06.2024).

**Hodel, Tobias, David Schoch, Christa**

**Schneider u. a.:** General Models for Handwritten Text Recognition: Feasibility and State-of-the Art. German Kurrent as an Example. In: *Journal of Open Humanities Data* 7 (2021).

**Hoffmann, Hartmut u. Rudolf Pokorny:**

Das Dekret des Bischofs Burchard von Worms. Textstufen – frühe Verbreitung – Vorlagen (*Monumenta Germaniae Historica Hilfsmittel* 12). München 1991.

**Holeczek, Heinz:** Erasmus von Rotterdam (1466/67–1536). Humanistische Profile – Erasmus im Profil. In: Paul G. Schmidt (Hg.): *Humanismus im deutschen Südwesten. Biographische Profile*. Sigmaringen 1993, S. 125–149.

**Jannidis, Fotis:** Zeichen und Zahlen. In: Ders., Hubertus Kohle u. Malte Rehbein (Hgg.): *Digital Humanities: Eine Einführung*. Stuttgart 2017, S. 59–67.

**Klaffki, Lisa, Stefan Schmunk u. Thomas**

**Stäcker:** Stand der Kulturgutdigitalisierung in Deutschland. Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums „Wissenschaftliche Sammlungen“. 2018 <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2018-26.pdf> (Zugriff: 08.06.2024).

**Kocher, Ursula:** Vom Nutzen der Hybridedition. In: *Editio* 33 (2019), S. 82–93.

**Krämer, Sybille:** Der ‚Stachel des Digitalen‘ – ein Anreiz zur Selbstreflexion in den Geisteswissenschaften? Ein philosophischer Kommentar zu den Digital Humanities in neun Thesen. In: *Digital Classics Online* 4, 1 (2018). <https://doi.org/10.11588/dco.2017.0.48490> (Zugriff: 08.06.2024).

**Krämer, Sybille:** Digitalism as a Cultural Technique: From Alphanumeric to AI. 2021. <https://www.goethe.de/prj/k40/en/eth/dig.html> (Zugriff: 08.06.2024).

**McKitterick, David:** *Old Books, New Technologies. The Representation, Conservation and Transformation of Books Since 1700*. Cambridge 2014.

**Mertgens, Andreas:** Ediarum: A Toolbox for Editors and Developers. In: *Ride* 11 (2019). <https://ride.i-d-e.de/issues/issue-11/ediarum/> (Zugriff: 08.06.2024).

**Muhlack, Ulrich:** Beatus Rhenanus (1485–1547). Vom Humanismus zur Philologie. In: Paul G. Schmidt (Hg.): *Humanismus im deutschen Südwesten*.

- Biographische Profile. Sigmaringen 1993, S. 195–220.
- Naveed, Humza, Asad Ullah Khan, Shi Qiu u. a.:** A Comprehensive Overview of Large Language Models. 2023. <https://doi.org/10.48550/arXiv.2307.06435> (Zugriff: 08.06.2024).
- Peter Robinson u. Elizabeth Solopova:** Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue. 1993. <https://zenodo.org/records/4050360> (Zugriff: 08.06.2024).
- Peter Shillingsburg:** The Literary Text in the Digital Age. In: Richard J. Finneran (Hg.): The Literary Text in the Digital Age (Editorial Theory and Literary Criticism). Ann Arbor MI 1996, S. 23–36.
- Pierazzo, Elena:** Digital Scholarly Editing. Theories, Models and Methods. New York 2015.
- Pinche, Ariane:** Generic HTR Models for Medieval Manuscripts. The CREM-MALab Project. In: Journal of Data Mining & Digital Humanities (2023). <https://doi.org/10.46298/jdmhdh.10252> (Zugriff: 08.06.2024).
- Raffel, Colin, Noam Shazeer, Adam Roberts u. a.:** Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. 2019. <https://doi.org/10.48550/arXiv.1910.10683> (Zugriff: 08.06.2024).
- Ralle, Inga H.:** Maschinenlesbar – menschenlesbar. In: Editio 30 (2016), S. 144–156.
- Rehbein, Malte:** Digitalisierung. In: Fotis Jannidis, Hubertus Kohle u. Malte Rehbein (Hgg.): Digital Humanities: Eine Einführung. Stuttgart 2017, S. 179–198.
- Romein, C. Annemieke, Tobias Hodel, Femke Gordijn u. a.:** Exploring Data Provenance in Handwritten Text Recognition Infrastructure: Sharing and Reusing Ground Truth Data, Referencing Models, and Acknowledging Contributions. Starting the Conversation on How We Could Get It Done. 2022. <https://zenodo.org/records/7267245> (Zugriff: 08.06.2024).
- Sahle, Patrick:** Digitale Editionsformen (Schriften des Instituts für Dokumentologie und Editorik 7). Norderstedt, Köln 2013a.
- Sahle, Patrick:** Digitale Editionsformen. Teil 2 (Schriften des Instituts für Dokumentologie und Editorik 8). Norderstedt 2013b.
- Schaßan, Torsten:** Wie können die Bibliotheken das Edieren diplomatischer Quelle unterstützen? In: Francia (2013), S. 299–303.
- Schonhardt, Michael:** Alter Wein in neue Schläuche? Einige Thesen zum Verhältnis von Wissenschaft und social media. 2014. <https://mittelalter.hypotheses.org/3975>. (Zugriff: 08.06.2024).
- Schonhardt, Michael:** bdd-segmentation-baselines. 2024a. <https://zenodo.org/records/10894508> (Zugriff: 08.06.2024).
- Schonhardt, Michael:** bdd-segmentation-data. 2024b. <https://zenodo.org/records/10882817> (Zugriff: 08.06.2024).
- Schonhardt, Michael:** bdd-segmentation-regions. 2024c. <https://zenodo.org/records/10890967> (Zugriff: 08.06.2024).
- Shevchuk, Ivan, Olivier Bonnerot, Christian Brockmann u. a.:** Wiedergewinnung verlorener Schrift und Analyse der Tintenzusammensetzung der beglaubigten Kopie der Handelsprivilegien Hamburgs, Threse B 15 a (von 14. Juli 1482), Bibliothek des Staatsarchives, Hamburg. 2021. <https://doi.org/10.25592/uhhfdm.9167> (Zugriff: 08.06.2024).
- Steding, Sören:** Warum noch drucken? Über die Nachteile gedruckter Editionen. 2001. <http://computerphilologie>.

- digital-humanities.de/jg01/steding.html  
(Zugriff: 08.06.2024).
- Terras, Melissa:** Digitization and Digital Resources in the Humanities. In: Claire Warwick, Melissa Terras u. Julianne Nyhan (Hgg.): *Digital Humanities in Practice*. Cambridge 2018, S. 47–70.
- Trovato, Paolo u. Michael D. Reeve:** Everything You Always Wanted to Know About Lachmann's Method. A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Structuralism, Gladistics, and Copy-Text (*Storie e linguaggi* 7). Padova 2014.
- Vander Meulen, David L. u. G. Thomas Tanselle:** A System of Manuscript Transcription. In: *Studies in Bibliography* 52 (1999), S. 201–212.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar u. a.:** Attention Is All You Need. 2017. <https://doi.org/10.48550/arXiv.1706.03762> (Zugriff: 08.06.2024).
- Wei, Jason, Xuezhi Wang, Dale Schuurmans u. a.:** Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. 2022. <https://doi.org/10.48550/arXiv.2201.11903> (Zugriff: 08.06.2024).
- Weigel, Harald:** „Nur was du nie gesehn wird ewig dauern“. Carl Lachmann und die Entstehung der wissenschaftlichen Edition (Rombach Wissenschaft Reihe Litterae). Freiburg 1989.
- West, Martin L.:** *Textual Criticism and Editorial Technique. Applicable to Greek and Latin Texts* (Teubner Studienbücher Philologie). Stuttgart 1973.
- Wilkinson, Mark, Michel Dumontier, IJsbrand Jan Aalbersberg u. a.:** The FAIR Guiding Principles for Scientific Data Management and Stewardship. In: *Scientific Data* 3 (2016). <https://doi.org/10.1038/sdata.2016.18> (Zugriff: 08.06.2024).
- Wolfram, Stephen:** What Is ChatGPT Doing ... and Why Does It Work? 2023. <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>. (Zugriff: 08.06.2024).
- Woolridge, Michael u. Nicholas Jennings:** Intelligent Agents: Theory and Practice. In: *The Knowledge Engineering Review* 10 (1995), S. 115–152.

# Interdisziplinär erforscht

## Das Phänomen der *Tagin* und *Otiyyot Meshunnot* als paläographische Besonderheit in mittelalterlichen Torarollen

### Kontakt

#### Dr. Dana Eichhorst,

Freie Universität Berlin,  
Institut für Judaistik,  
Fabeckstraße 23–25, 14195 Berlin,  
dana.eichhorst@fu-berlin.de  
 <https://orcid.org/0000-0002-4085-3337>

#### Laura Frank,

Karlsruher Institut für Technologie (KIT),  
Scientific Computing Center (SCC),  
Hermann-von-Helmholtz-Platz 1,  
76344 Eggenstein-Leopoldshafen,  
laura.frank@kit.edu  
 <https://orcid.org/0000-0001-6286-2771>

#### Dr. Danah Tonne,

Karlsruher Institut für Technologie (KIT),  
Scientific Computing Center (SCC),  
Hermann-von-Helmholtz-Platz 1,  
76344 Eggenstein-Leopoldshafen,  
danah.tonne@kit.edu  
 <https://orcid.org/0000-0001-6296-7282>

#### Katharina H. Wendl, Freie

Universität Berlin, Institut für Judaistik,  
Fabeckstraße 23–25, 14195 Berlin,  
katharina.wendl@fu-berlin.de  
 <https://orcid.org/0009-0001-2664-2453>

**Abstract** In Judaism, Torah scrolls must be produced and written according to strict traditional rules. Some letters of the sacred text have distinctive decorations: the so-called *tagin* ('little crowns') and *otiyyot meshunnot* ('special letters'). However, despite the binding guidelines for writing a Torah scroll, it is evident that the design and shape of these little crowns and special letters differ greatly in various surviving Torah scrolls and manuscripts. This article shows how the research project 'ToRoll: Materialized Holiness' uses digital methods to examine *tagin* and *otiyyot meshunnot* as palaeographic and metaphysical peculiarities in Torah scrolls to comprehensively present the phenomenon of *tagin* and *otiyyot meshunnot* for the first time. In addition to the history of the studied texts and the at times complex manuscript situation, the diverse nature of the various sources poses a particular challenge. We use critical digital editions and web annotations to meet the requirements of these complex text structures. This enables us to link variants of the letter decorations in Torah scrolls with the accompanying commentary literature. The use of diverse digital methods and principles allows us to grasp the phenomenon of *tagin* and *otiyyot meshunnot* on a philological, palaeographical, and historical level. Combining expertise and methodologies of different disciplines is both a challenge and an opportunity. This article is intended to highlight specific solutions for the research project and open up perspectives for other disciplines.

**Keywords** Jewish Scribal Traditions; Digital Medieval Studies; TEI-XML; Torah Scrolls; Web Annotation Data Model

## 1 Einleitung

Im Judentum sind Torarollen für den rituellen und liturgischen Gebrauch in der Synagoge essentiell. Da der Text der Tora als von Gott offenbart und folglich als heilig erachtet wird, ist jeder einzelne Buchstabe und jedes Zeichen von Bedeutung.<sup>1</sup> Aus diesem Grund hat sich im Laufe der Zeit eine bis heute praktizierte Schreibtradition und parallel dazu eine umfangreiche Schreiberliteratur entwickelt, in der diskutiert, erklärt und letztlich festgelegt wird, nach welchen Regeln Torarollen angefertigt und geschrieben werden sollen. Mittelalterliche Torarollen bezeugen jedoch ein seltsames Phänomen: Manche Buchstaben des heiligen Textes zeichnen sich in ihrer Ausgestaltung durch die sogenannten *Tagin* („Krönchen“) oder durch eine besondere Form, der sie auch ihren Namen *Otijjot Meshunnot*, ‚besondere Buchstaben‘, verdanken, aus. *Tagin* sind Ergänzungen am Buchstaben in Form von kleinen Strichen, die in der jüdischen Literatur mitunter auch als ‚kleine Dornen‘ bezeichnet werden. Die *Otijjot Meshunnot* hingegen sind mit Kringeln, Schnörkeln bzw. ‚Fähnchen‘ ausgeführt oder weisen eine besonders gewundene Linienführung auf.

Diese besondere, teils äußerst dekorative Schreibweise der hebräischen Buchstaben der Heiligen Schrift mit *Tagin* oder als *Otijjot Meshunnot* verändert weder die Syntax des biblischen Textes noch beeinflusst sie seine Semantik. Das bedeutet, dass der Text in den Torarollen keine inhaltliche Veränderung erfährt, da der Buchstabe trotz der *Tagin* bzw. *Otijjot Meshunnot* selbst lesbar bleibt. *Tagin* und *Otijjot Meshunnot* sind keine diakritischen Zeichen oder masoretischen Anmerkungen,<sup>2</sup> wie sie in Codices und Büchern der hebräischen Bibel zu finden sind. Sie haben zweifellos dekorativen Charakter, jedoch ist anzunehmen, dass sich ihre Bedeutung nicht nur auf die Ästhetik beschränkt. Es gibt vielmehr Grund zur Annahme, dass sie auf verschiedene Ebenen bzw. Motive<sup>3</sup> der jüdischen Tradition verweisen, die über den Bibeltext hinausgehen. Auch wenn bisher noch nicht hinreichend erforscht ist, nach welchen Prinzipien und Regeln mittelalterliche Torarollen mit *Tagin* und *Otijjot Meshunnot* versehen wurden, kann doch mit Sicherheit davon ausgegangen werden, dass sie als wesentlicher Bestandteil einer Torarolle erachtet wurden.

---

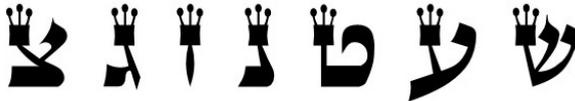
1 Vgl. Stern 2017; Martini 2022, S. 14–26; Fishbane 2016, S. 211–250.

2 Unter ‚masoretische Anmerkungen‘ fallen die Vokalisierung und Akzentuierung des hebräischen Textes, dessen Einteilung in kleinere Leseabschnitte und die Auflistung von Lesarten, die vom überlieferten Text abweichen.

3 Ein Forschungsansatz versteht die *Tagin* etwa als Ausdruck einer exegetischen Tradition; andere vermuten in ihnen gewisse ‚Marker‘, die beim Lesen des Toratextes während der Liturgie eine Rolle spielen.

## 1.1 Rabbinischer Anspruch und Schreibpraxis im Widerspruch

Seit der Antike haben jüdische Autoritäten Regeln und Anweisungen zum Schreiben von Torarollen formuliert und dabei auch *Tagin* berücksichtigt. Eine quasi obligat gewordene Anweisung der Rabbinen ist im babylonischen Talmud überliefert,<sup>4</sup> wonach sieben Buchstaben des hebräischen Alphabets mit jeweils drei *Tagin* versehen werden sollen (vgl. Abb. 1).<sup>5</sup>



**Abb. 1** | Sieben Buchstaben, die gemäß der rabbinischen Vorgabe mit *Tagin* geschrieben werden. Graphik: ‚ToRoll: Materialisierte Heiligkeit‘. CC BY-NC-ND.

Diese konkreten rabbinischen Vorgaben zum Schreiben der Buchstaben in einer Torarolle mit oder ohne *Tagin* erlauben allerdings keinen Rückschluss auf den Ursprung, die Funktion oder die Bedeutung der *Tagin*. Die Entwicklung der *Tagin* und *Otillot Meshunnot* ist bisher ebenso unzureichend erforscht wie jene der jüdischen rituellen Schreibpraxis insgesamt und des umfangreichen Diskurses darüber.

4 Traktat Menahot 29b (bT Menahot 29b). Die einzige deutsche Talmudausgabe basiert auf der Übersetzung von Goldschmidt 1897–1935. Die hebräische Standardedition des babylonischen Talmuds hingegen basiert auf der Ausgabe Wilna 1880–1886.

5 Das (Nicht-)Vorhandensein von *Tagin* und *Otillot Meshunnot* in einer Torarolle hat aus halachischer (jüdisch-legalistischer) Perspektive keinen Einfluss auf ihre Validität. A priori sollen die erwähnten sieben Buchstaben (*shin*, *ayin*, *tet*, *nun*, *zayin*, *gimmel*, *sade*) mit drei *Tagin* verziert werden (bT Menahot 29b). Werden diese von einem Schreiber überhaupt nicht, unvollständig, fehlerhaft oder von der Norm abweichend ausgeführt, ist die Torarolle dennoch für den rituellen Gebrauch geeignet; vgl. Karo: Shulḥan Arukh, Oraḥ Ḥajjim, 36:3; Maimonides: Mishneh Torah, Hilkhot Tefillin, Mezuzah ve-Sefer Torah, 7:8–9. Nur eine Minderheit erachtet eine Torarolle als ungeeignet, sollten *Tagin* fehlen; vgl. ‚Shimusha Rabbah‘ in Jaakov ben Asher: Arbaah Ṭurim, Oraḥ Ḥajjim, 36:5. Daneben gibt es sechs weitere Buchstaben (*bet*, *dalet*, *qof*, *het*, *yod*, *he*), die einen *Tag* aufweisen sollen. Ein Fehlen dieses *Tag* hat aber nach Meinung aller keine Auswirkungen auf die Validität einer Torarolle. Mit der Zeit ist es praktisch obligatorisch geworden, die erwähnten sieben Buchstaben mit drei *Tagin* und die übrigen sechs mit einem *Tag* zu dekorieren; vgl. Ganzfried: Keset ha-Sofer, 5:4, Kagan: Mishnah Berurah, 36:15. Da das talmudische Diktum *Tagin* erlaubt (ihre Verwendung aber nicht auf die genannten Buchstaben beschränkt), sahen Schreiber in den *Tagin* und *Otillot Meshunnot* offenbar eine Möglichkeit, ihrer Kreativität beim Schreiben freien Lauf zu lassen.

## 1.2 Zeugen eines Regelbruchs

Neben der Frage nach der grundlegenden Bedeutung der *Tagin* und *Otíjjot Meshunnot* geben insbesondere die mittelalterlichen Torarollen weitere Rätsel auf. So ist bei vielen Rollen entgegen der rabbinischen Vorgabe die Auszeichnung der Buchstaben mit *Tagin* sehr unregelmäßig. Einige Rollen weisen lediglich die im Talmud vorgegebenen drei Striche auf den oben genannten sieben Buchstaben auf, andere hingegen zeigen eine große Fülle an ‚Krönungs-Varianten‘. In den von uns untersuchten Torarollen lassen sich auf allen 22 Buchstaben des hebräischen Alphabets ein bis neun *Tagin* finden (vgl. Abb. 2). Eine solche Vielzahl an *Tagin* und Vielfalt an *Otíjjot Meshunnot* wird in der klassischen rabbinischen Literatur jedoch thematisch nur wenig berührt.

Seit dem 10. Jahrhundert werden *Tagin* und *Otíjjot Meshunnot* zunehmend zum Gegenstand innerjüdischer Diskurse. Im Verlauf der Jahrhunderte hat sich im Judentum eine umfangreiche Auslegungstradition entwickelt, die den ‚Verzierungen‘ unterschiedliche Bedeutungen zuschreibt und die in einer Vielzahl von verschiedenen Quellen überliefert ist. Diese umfassen religionsgesetzliche<sup>6</sup> und



**Abb. 2 |** Buchstaben mit *Tagin* und verschiedene *Otíjjot Meshunnot* aus Ms. or. fol. 1215. Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Orientabteilung. <http://resolver.staatsbibliothek-berlin.de/SBB00005AA700000000>. Erstellt mit <https://yasgui.org/>.

<sup>6</sup> Zwei der bekanntesten religionsgesetzlichen Regelwerke sind ‚Mishneh Torah‘ (1170–1180) von Moses Maimonides und ‚Shulḥan Arukh‘ (1565) von R. Joseph Karo. Diese wurden zahlreich kommentiert und adaptiert. Einige halachische Werke widmen sich ausschließlich bestimmten Themen, etwa der jüdischen Schreibtradition, wie z. B. ‚Bnei Jonah‘ (1803) von Jonah Landsofer und ‚Keset ha-Sofer‘ (1834/1871) von Shlomo Ganzfried.

exegetische Texte<sup>7</sup> sowie praktische Schreiberhandbücher.<sup>8</sup> Ein Standardisierungsprozess, der die *Tagin* und *Otillot Meshunnot* berücksichtigt, scheint unseren Erkenntnissen zufolge erst mit dem ausgehenden Mittelalter einzusetzen.

## 2 Stand der Forschung

Die vergleichsweise junge Forschung zu Torarollen fokussierte sich bislang in erster Linie auf die materialen Aspekte wie beispielsweise die Beschaffenheit der Schreibhäute und Tinten sowie auf Merkmale des Layouts.<sup>9</sup> Ein Forschungsdesiderat hingegen stellt die metaphysische Bedeutung der *Tagin* und *Otillot Meshunnot* jenseits der Materialforschung dar. Die Inkonsistenzen in Bezug auf die Ausgestaltung der Buchstaben, wie sie in den Torarollen zu beobachten sind, lassen sich weder allein durch die Torarollen selbst noch durch die Vorgaben in den überlieferten Schreiberhandbüchern erklären. Daher ist es erforderlich, die Analyse der materialen Aspekte von Torarollen mit einer multiperspektivischen Untersuchung zu verbinden, in der *Tagin* und *Otillot Meshunnot* als paläographisches wie auch als religions-, kultur- und ideengeschichtliches Phänomen betrachtet werden. In diesem Kontext ist auch die bereits erwähnte jüdische Auslegungstradition von Belang. Ihre Auswertung unter Berücksichtigung von paläographischen und materialanalytischen Befunden liefert neue Erkenntnisse über die Geschichte, Funktion und Bedeutung der *Tagin* und *Otillot Meshunnot*.

### 2.1 Entwicklung interdisziplinärer Forschungsansätze

Die strukturierte Sammlung aller verfügbaren Informationen zu erhaltenen mittelalterlichen Torarollen sowie die Einbindung des Textkorpus der Auslegungstradition und der Schreiberliteratur in einem umfangreichen Wissensspeicher ermöglichen erstmals eine umfassende Erschließung dieses Forschungsgegenstands aus unterschiedlichen Perspektiven und durch verschiedene Disziplinen. Dies erlaubt eine Forschung, die über die Methoden der jeweiligen Einzeldisziplinen der Paläographie, Materialforschung sowie Philologie und Geschichtswissenschaft

7 Z. B. ‚Midrasch Rabbi Akiva al ha-Tagin‘ und der Bibelkommentar ‚Baal ha-Ṭurim‘ des R. Jaakov ben Asher (ca. 1270–1340).

8 Über die Jahrhunderte sind jüdische rituelle Schreiberhandbücher in verschiedenen kulturellen Kontexten entstanden. Eines der bedeutendsten mittelalterlichen Handbücher über die Herstellung von Tefillin ist der Shimshon ben Eliezer zugeschriebene ‚Baruch she-Amar‘. Für die Neuzeit ist das Handbuch ‚Keset ha-Sofer‘ von Ganzfried zu erwähnen.

9 Vgl. Del Barco 2020, S. 91–118; Dunkelgrün 2021, S. 47–92; Nehring, Gordon u. Rabin 2021; Nehring, Gordon u. Rabin 2022; Cruz 2023.

hinausgeht. Der Rückgriff auf digitale Werkzeuge und Methoden führt zudem zu neuen Forschungsfragen und -ansätzen.

Eine solche Forschungsfrage betrifft beispielsweise die Terminologie der *Tagin* und *Otillot Meshunnot*. Die Erstellung digitaler Editionen von Texten aus dem Korpus der Kommentar- und Schreiberliteratur führte durch die Erfassung und Auswertung aller relevanten Termini zu der Erkenntnis, dass innerhalb der jüdischen Tradition keine einheitliche Terminologie existiert. Dies begründet in Folge das Desiderat einer konsistenten Fachterminologie des Forschungsfeldes.

## 2.2 Digitale und geisteswissenschaftliche Methoden im Verbund

Bislang wurden die forschungsrelevanten Quellen nur unzureichend erschlossen und kritisch ediert. Viele der berücksichtigten Texte sind in zahlreichen voneinander abweichenden Versionen und in verschiedenen Manuskriptzeugen überliefert. Andere weisen intertextuelle Bezüge zu anderen Textquellen auf. Die heterogene Handschriftenlage wäre mit einer klassischen Edition, selbst mit einer synoptischen, nur schwer adäquat zu bewältigen. Herkömmliche analoge Editionen vermögen zudem nur begrenzt, inhaltliche und historische Bezüge zwischen den verschiedenen Texten darzustellen. Die Schnittstelle von digitalen und klassischen philologischen Methoden erweitert die bisherigen Mittel der Texterschließung. Der klassische lineare Zugriff auf einen Text kann durch die Verwendung und Weiterentwicklung digitaler Werkzeuge um einen mehrdimensionalen Zugriff und eine alineare Lesbarkeit ergänzt werden. Im Folgenden soll anhand erster Forschungsergebnisse des Verbundprojekts ‚ToRoll: Materialisierte Heiligkeit‘ gezeigt werden, wie digitale, geistes- und hilfswissenschaftliche Methoden und Arbeitsweisen ineinandergreifen und zu neuen Fragestellungen und Antworten führen können, die Transferpotential für andere Fächer haben.

## 3 Methodik

Um das Phänomen der *Tagin* und *Otillot Meshunnot* zu erfassen, werden zwei verschiedene Gruppen von Quellen berücksichtigt, die hauptsächlich in Manuskripten erhalten sind: zum einen mittelalterliche Torarollen und zum anderen ein umfangreiches Textkorpus der jüdischen Auslegungs- und Schreiberliteratur. Ein wesentliches Merkmal des Toratextes ist, dass er als heilig und unveränderlich gilt. Diese Textkonstanz trifft aber nicht auf die ebenfalls untersuchten Texte der jüdischen Auslegungs- und Schreiberliteratur zu, die zum Teil sehr heterogen und fluide sind.

Der Forschungsansatz bedient sich der Expertise und Methodik verschiedener Disziplinen: Zur Beantwortung der Frage, wie und warum die *Tagin* und *Otijjot Meshunnot* entstanden sind, wird das Textkorpus der Schreiber- und Kommentarliteratur in erster Linie mit den klassischen philologischen Mitteln der Geschichtswissenschaft untersucht. Die Paläographie gibt Aufschluss über die Gestaltungsform der *Tagin* und *Otijjot Meshunnot* und ihre Entwicklung. Die Informationswissenschaften ermöglichen eine umfänglichere Erfassung, Anreicherung und Auswertung aller gewonnenen Erkenntnisse in einer eigens geschaffenen digitalen Infrastruktur, ohne die die Analyse von *Tagin* und *Otijjot Meshunnot* deutlich mühsamer wäre.<sup>10</sup> Die Infrastruktur besteht aus einem Forschungsdatenrepositorium, einem Vokabulardienst, einem Annotationsdienst sowie verschiedenen Diensten zur Datenanalyse und -visualisierung. Die dafür entwickelte ‚virtuelle Torarolle‘ erlaubt es, eine Vielzahl an disziplinübergreifenden Daten, Metadaten und Annotationen bezüglich der *Tagin* und *Otijjot Meshunnot* zu akkumulieren, zu verknüpfen und gezielt zu visualisieren, wodurch zudem ein direkter Vergleich zwischen verschiedenen Torarollen und wichtigen Quellen der Schreiberliteratur möglich wird.

Anhand eines konkreten Fallbeispiels soll gezeigt werden, welche Analysemöglichkeiten die ‚virtuelle Torarolle‘ bietet und wie die Verknüpfung der einzelnen Fachdisziplinen dazu beiträgt, die Geschichte, Entwicklung und Bedeutung der *Tagin* und *Otijjot Meshunnot* umfassend zu erforschen. Auch wenn im Folgenden ein ausgewählter Buchstabe und seine Ausgestaltung im Fokus der Betrachtung stehen, ist die vorgestellte Methodik uneingeschränkt auf andere Buchstaben oder auch Mengen von Buchstaben anwendbar.

### 3.1 Dem Regelbruch auf der Spur

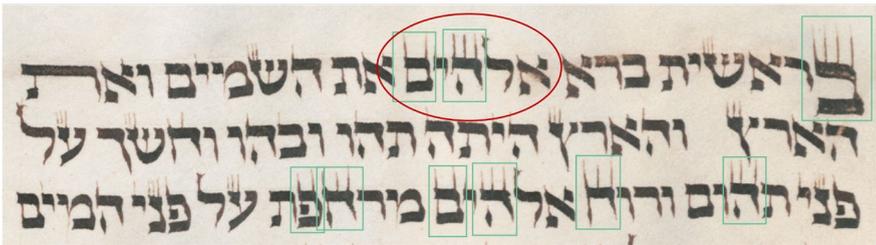
Vor uns liegt eine Torarolle, die in der Staatsbibliothek zu Berlin, Preußischer Kulturbesitz aufbewahrt wird (Ms. or. fol. 1215) und deren Entstehung aufgrund ihrer paläographischen Charakteristika und einer Analyse des Schreibstoffs auf die Zeit zwischen 1300 und 1400 datiert wird. Die Rolle, einseitig auf Pergament beschrieben, besteht aus 86 einzelnen miteinander vernähten Folios; sie misst

---

<sup>10</sup> An dieser Stelle sei die jahrzehntelange akribische Forschungsarbeit von Franz Hubmann und Josef M. Oesch zu Torarollen und darin enthaltenen paläographischen Sonderbarkeiten erwähnt. In einer Vielzahl von Tabellen und Listen haben sie quantitative Ergebnisse ihrer sorgfältigen Studien mit Bleistift auf Papierbögen zusammengetragen und festgehalten. Ihre Arbeit stellt einen wertvollen Beitrag für die Forschung in diesem Bereich dar. Für die freundliche Überlassung ihrer handschriftlichen, nicht veröffentlichten Unterlagen und Analysen gebührt ihnen unser aufrichtiger Dank. Publierte Ergebnisse finden sich u. a. in Hubmann u. Oesch 2012.

in ihrer Länge ca. 50 m bei einer Höhe von ca. 80 cm. Man kann sie schwerlich handlich nennen. Sie lässt sich auch nicht wie ein Codex oder ein Buch einfach durchblättern, sondern nur mit relativ viel Muskelkraft auf- und abrollen. Möchte man nun einen bestimmten Abschnitt des Toratextes lesen, schlägt man nicht etwa die entsprechende Seite auf – einen Index oder Seitenzahlen gibt es traditionell nicht –, sondern man rollt und rollt, wobei einem der Staub von Jahrhunderten um die Nase weht. Währenddessen wird man einer Vielzahl von *Tagin* gewahr. Augenscheinlich sind in dieser Torarolle nicht nur die ‚erlaubten‘ sieben Buchstaben mit den obligatorischen drei kleinen Strichlein versehen, sondern auch viele andere Buchstaben mit mehr oder weniger als den vorgeschriebenen drei *Tagin*. Gleich zu Beginn des Toratextes, auf dem dritten Wort im Abschnitt Gen 1:1, trägt zum Beispiel der Buchstabe *he* im Gottesnamen *Elohim* (אלהים) überraschenderweise drei *Tagin*, und auch der letzte Buchstabe dieses Wortes, *mem*, trägt – ebenfalls regelwidrig – *Tagin*, allerdings nur zwei (vgl. Abb. 3). Wie ist dies zu erklären?

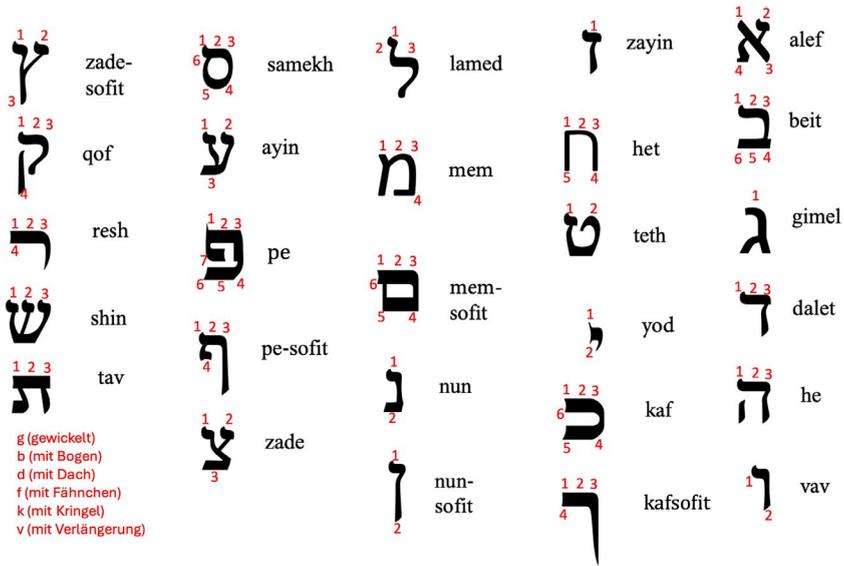
Hätte man viel Zeit und Platz, könnte man die Torarolle in ihrer Gesamtlänge ausrollen und mit einer Analyse des Textes beginnen. Allerdings würde man tunlichst vermeiden, das alte Manuskript durch Hinzufügen von Notizen oder Ähnlichem zu beschädigen. Was tun?



**Abb. 3** | Vers Gen 1:1–2 in der Torarolle Ms. or. fol. 1215 mit regelkonformen und regelwidrigen *Tagin* (grün umrandet) wie beispielsweise im rot hervorgehobenen Wort *Elohim* auf dem dritten Wort von rechts. Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Orientabteilung.

### 3.2 Das Erfassen und Bündeln von Regelwidrigkeiten

Zunächst stellt sich die Frage, wie sich die unterschiedlichen Erscheinungsformen der *Tagin* und *Otillot Meshunnot* der Torarolle dokumentieren und vergleichen lassen. Eine manuelle Auszeichnung erfordert klare Definitionen und Beschreibungen, während eine automatische Auswertung und Visualisierung standardisierte, maschinenlesbare Begriffe erfordert. Zu diesem Zweck wurde in einem



**Abb. 4 |** Für jeden Buchstaben sind die Positionen, an denen Verzierungen wie Krönchen, Kringlel oder Fahnen vorkommen können, durchnummeriert, was die Unterscheidung sowie die eindeutige Referenzierung der einzelnen regelwidrigen Dekorationen erlaubt. Für das betrachtete *he* sind drei Positionen vorgesehen. Graphik: ‚ToRoll: Materialisierte Heiligkeit‘. CC BY-NC-ND.

ersten Schritt mithilfe des Vokabulardienstes Evoks<sup>11</sup> ein hierarchisches Vokabular angelegt (vgl. Abb. 4 für das Schema zur Variantenbenennung), das zurzeit über 300 Begriffe umfasst und laufend erweitert wird. Allein für den Buchstaben *he* sind zehn ‚Krönungsvarianten‘ verzeichnet.

Die hochauflösenden Bilddigitalisate sowie die Textrepräsentation (das heißt die fünf Bücher Mose) der Torarolle (Ms. or. fol. 1215) werden in ein digitales Forschungsdatenrepositorium überführt und als strukturiertes Torarollenobjekt nachhaltig nutzbar zur Verfügung gestellt. Neben inhaltlichen Metadaten zum Original wie zum Beispiel der Signatur, der Angabe der besitzenden Institution und einer Objektbeschreibung werden auch administrative Metadaten der digitalen Repräsentation gespeichert, wie ein eindeutiger Identifier oder das Erstellungsdatum. Zusätzlich werden Verknüpfungen zu Textobjekten mit einer digitalen Repräsentation des Torarollentextes und zu Bildobjekten mit Digitalisaten der Torarolle realisiert. Eine solche Struktur mit standardisierten Formaten ist unverzichtbar

<sup>11</sup> Vgl. Kraus 2024. Der Vokabulardienst erlaubt die Erstellung, Bearbeitung, Speicherung und Publikation von Vokabularen, Taxonomien und Begriffslisten basierend auf SKOS (*Simple Knowledge Organization System*); vgl. Miles u. Bechhofer 2009.

für eine langfristige Speicherung und ermöglicht die Nutzung von Diensten zur weiteren Anreicherung und Analyse.<sup>12</sup>

Für die Untersuchung der *Tagin* und *Otijjot Meshunnot* in der Torarolle Ms. or. fol. 1215 werden alle Buchstaben mit regelwidrigen *Tagin* und alle ‚besonderen Buchstaben‘ erfasst. Dies geschieht sowohl im Bild als auch an der passenden Stelle im Textdigitalisat als digitale Webannotation.<sup>13</sup> Eine passgenaue Modellierung der fachwissenschaftlichen Anforderungen ist dabei ohne Probleme möglich. Die Annotationen der Torarolle enthalten somit Informationen über das zu annotierende Objekt (*target*). Dieses kann entweder ein Ausschnitt einer Bildressource (svg-Selektor) oder eine konkrete Stelle in einer Textressource (xml-Selektor) sein. Angaben über die Art der Buchstabenvariante und die Klassifizierung der *Tagin* und *Otijjot Meshunnot* gemäß dem Vokabular werden als Annotationsinhalte (*bodies*) abgespeichert.<sup>14</sup> Zur Erstellung der Annotationen wird ein Annotationsdienst<sup>15</sup> verwendet, der eine graphische Oberfläche als Schnittstelle zu den Forschenden bereitstellt und zusätzliche Filter- und Suchmöglichkeiten bietet.

Folio für Folio werden alle regelwidrigen *Tagin* und *Otijjot Meshunnot* in Text und Bild digital annotiert. Der angeschlossene Annotationspeicher<sup>16</sup> erlaubt flexible SPARQL-Abfragen<sup>17</sup> und damit dynamische Zugänge zu den erfassten Informationen. So sind sehr komplexe Abfragen, jeweils passend zur spezifischen

---

12 Verschiedene technische Komponenten gewährleisten dabei Konformität der Daten mit den sogenannten FAIR-Prinzipien (vgl. Wilkinson, Dumontier, Aalbersberg u. a. 2016), d. h. die Forschungsdaten sollen auffindbar (*Findable*), zugänglich (*Accessible*), interoperabel (*Interoperable*) und wiederverwendbar (*Reusable*) sein. Das Repository basiert technologisch auf Entwicklungen des KIT Data Managers; vgl. Jejkal u. Hartmann 2020. Dieses generische Framework erlaubt die Implementierung von Forschungsdatenrepositorien in verschiedensten Disziplinen. Durch die Nutzung standardisierter Modelle und Formate können Anwendungsfälle von beispielsweise den Materialwissenschaften bis zu den digitalen Geisteswissenschaften abgedeckt werden.

13 An dieser Stelle sei ein großes Dankeschön an Juan E. Mora, Dr. Emese Kozma, Aram Abu-Saleh und Konstantin Paul für ihre detaillierte Annotationsarbeit der in diesem Artikel diskutierten Torarollen und Manuskripte ausgesprochen.

Eine digitale Webannotation ist die Anreicherung von Webressourcen durch zusätzliche Informationen, z. B. durch Webressourcen oder textuellen Inhalt. Seit Februar 2017 existiert mit der W3C-Empfehlung ‚Web Annotation Data Model‘ (WADM; vgl. Young, Ciccarese u. Sanderson 2017) ein leistungsfähiges Modell, das die Nutzbarkeit und Austauschbarkeit zwischen verschiedenen Systemen unterstützen soll.

14 Allgemein sind verschiedene Inhalte wie Bild, Text, Audio oder Video sowohl für das *target* als auch für *bodies* möglich. WADM basiert auf maschinenlesbaren *Semantic Web*-Standards, stellt aber auch eine menschenlesbare JSON LD (*JavaScript Object Notation for Linked Data*)-Repräsentation der Annotation zur Verfügung.

15 Vgl. Tonne, Götzelmann, Pütz u. a. 2024.

16 Entwickelt gemäß W3C-Empfehlung ‚Web Annotation Protocol‘ mit RDF (*Resource Description Framework*)-Backend; vgl. Sanderson 2017.

17 Vgl. Harris u. Seaborne 2013.

Forschungsfrage, und darauf aufbauende Analysen und vielfältige Visualisierungen möglich. Durch Ausnutzung der Hierarchie der Vokabulartermini kann beispielsweise gezielt nach allen Dekorationsvarianten eines Buchstabens oder allen Varianten eines bestimmten Phänomens (zum Beispiel ‚gewickelt‘ oder ‚4 Krönchen‘) gesucht werden. Suchen wir exemplarisch in den beiden ersten Kapiteln (Gen 1–2) nach unserem Buchstaben *he*, um zu prüfen, ob der Buchstabe eventuell nur in diesem einen Fall in Gen 1:1 mit drei *Tagin* geschrieben wurde, ergibt sich folgendes Bild (Tab. 1):

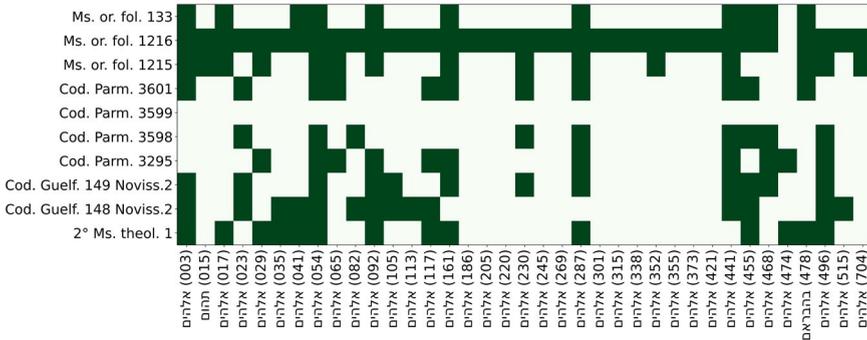
**Tabelle 1** | Suchergebnisse regelwidrig dekorierte *he* in Gen 1–2. Zur Identifikation der konkreten Stelle wurden die Worte aufsteigend nummeriert (WordID); zusätzlich wird die Klassifikation der Buchstabenvariante gemäß dem vorgestellten Vokabular angegeben.

<b>WordID</b>	003	015	017	029	054	065	092	161
<b>Wort</b>	אלהים	תהום	אלהים	אלהים	אלהים	אלהים	אלהים	אלהים
<b>Klassifikation</b>	he111							
<b>WordID</b>	230	287	352	441	478	496	704	
<b>Wort</b>	אלהים							
<b>Klassifikation</b>	he111							

### 3.3 Quantitative und qualitative Analyse gekrönter Buchstaben

Es zeigt sich, dass ein Schreiber in dieser Torarolle den Buchstaben *he* nicht nur in diesem einen Bibelvers, sondern an insgesamt 15 Stellen in den ersten zwei Genesis-Kapiteln mit jeweils drei *Tagin* versehen hat, davon 14-mal im Gottesnamen *Elohim* und ein anderes Mal im hebräischen Wort für ‚Abgrund‘ (*tehom* – תהום, WordID 015, Gen 1:2). Auch an dieser Stelle hat der Schreiber den Buchstaben *he* mit drei *Tagin* versehen (vgl. Abb. 3).

Wir wissen nicht, wieso er das tat. Aus den 14 Fällen, in denen der Schreiber den Gottesnamen ‚regelwidrig‘ mit *Tagin* verziert hat, könnte man schließen, dass sein Ansinnen vielleicht darin bestand, den Gottesnamen auf diese Weise hervorzuheben. Die weitere quantitative Auswertung ergibt Folgendes: Das Wort *Elohim* kommt in den ersten beiden Kapiteln von Genesis insgesamt 46-mal vor. Allerdings findet sich an diesen Stellen nur ein *Tag* auf dem *he* oder auch gar keine *Tagin*. Das Ergebnis wirft einige Fragen auf: Weshalb hat der Schreiber nicht in jedem Fall den Buchstaben *he* im Gottesnamen *Elohim* mit *Tagin* geschrieben? Wieso versieht er darüber hinaus das *he* in Gen 1:2 im Wort für ‚Abgrund‘ mit *Tagin*? Finden sich auch in anderen Torarollen diese drei *Tagin* oder weitere Kombinationen auf dem Buchstaben *he*? Haben wir es hier mit einer Schreibtradition



**Abb. 5 |** Vorkommen von *Tagin* auf dem *he* in Gen 1–2 mit zehn Torarollen im Vergleich. Die y-Achse erfasst die Signaturen der Torarollen, die x-Achse Wörter (inklusive WortID) mit *Tagin*. Markiert wird ein Feld, wenn die Torarolle ein Wort mit *Tagin* auf dem *he* aufweist. Wörter ohne *Tagin* sind für eine übersichtlichere Darstellung ausgelassen. Graphik: ‚ToRoll: Materialisierte Heiligkeit‘. CC BY-NC-ND.

zu tun, die die Schreibregeln der Rabbinen, wie sie im Talmud festgehalten sind, um weitere Regeln – hier das *he* betreffend – ergänzt hat?

Die Analyse (vgl. Abb. 5) zeigt, dass sich die zehn untersuchten Torarollen in Bezug auf die Dekoration des *he* in den Kapiteln 1 und 2 stark voneinander unterscheiden. Keine Rollen sind identisch miteinander. Während der Schreiber der Torarolle Ms. or. fol. 1216 etwa in den ersten zwei Kapiteln besonders viele *he* mit mehreren *Tagin* versehen hat, sind Schreiber anderer Rollen, etwa Ms. or. fol. 133, viel spärlicher mit Tinte und Feder vorgegangen. Cod. Parm. 3599 weist überhaupt keine *Tagin* auf dem Buchstaben *he* auf, weicht im Gegensatz zu anderen analysierten Torarollen also nicht von den im Talmud formulierten Regeln zur Buchstabendekoration ab. Eine nähere Betrachtung der übrigen Rollen ergibt, dass manche Gottesnamen besonders häufig mit *Tagin* ausgezeichnet werden (zum Beispiel WortID 054, 287 und 441 in jeweils acht bzw. neun von zehn Rollen). Es scheint also unter rituellen Schreibern im Mittelalter eine weitverbreitete Tendenz gegeben zu haben, den Buchstaben *he* mit *Tagin* zu versehen. Nach welchen Vorgaben haben sich die Schreiber dabei gerichtet?

### 3.4 Eigenständige Schreibtraditionen im Mittelalter

An dieser Stelle soll auf ein Werk verwiesen werden, das im Zusammenhang mit der Erforschung der *Tagin* und *Otillot Meshunnot* von Bedeutung ist: der sogenannte ‚Sefer *Tagin*‘ (‚Buch der *Tagin*‘). Herkunft und Urheberschaft des Textes sind bislang noch ungeklärt; er ist jedoch mit hinreichender Sicherheit

ins Frühmittelalter zu datieren.<sup>18</sup> Das kleine Büchlein bietet listenhaft Auskunft darüber, welche Buchstaben in den Versen einer Torarolle mit welcher Anzahl an *Tagin* und welche Buchstaben als *Otijjot Meshunnot* geschrieben werden sollen. Die Anweisungen umfassen auch Buchstaben, die nicht zu den sieben im Talmud genannten gehören. Es wird daher angenommen, dass der ‚Sefer Tagin‘ als eine Vorlage oder Handreichung für mittelalterliche Torarollenschreiber gedient haben könnte. Allerdings weisen die erhaltenen Manuskriptzeugen des ‚Sefer Tagin‘ sowohl in inhaltlicher als auch struktureller Hinsicht signifikante Unterschiede auf.<sup>19</sup> Es ist denkbar, dass diese divergierenden Versionen (regional) verschiedene Schreibtraditionen abbilden, was sich in der Heterogenität der erhaltenen mittelalterlichen Torarollen widerzuspiegeln scheint.

Da es bislang keine kritische Edition des ‚Sefer Tagin‘ gibt, die alle erhaltenen Versionen des Textes einbezieht, wird diese Forschungslücke nun mittels einer synoptischen digitalen Edition aller Textzeugen geschlossen. Alle bekannten Manuskriptzeugen des ‚Sefer Tagin‘ sind transkribiert und mittels TEI-Richtlinien<sup>20</sup> in eine digitale Edition überführt worden. Analog zu einem Torarollenobjekt werden sie als Manuskriptobjekte im Forschungsdatenrepositorium angelegt. Zusätzlich werden passende Torarollenobjekte angelegt, die gemäß den jeweiligen Anweisungen des ‚Sefer Tagin‘ annotiert werden. Im Fall unseres Buchstabens *he* ist es denkbar, dass sich im ‚Sefer Tagin‘ konkrete Vorgaben zu diesem Buchstaben finden. Eine Suchanfrage und ein Abgleich von allen im Repositorium angelegten Versionen des ‚Sefer Tagin‘ mit der annotierten Torarolle liefert, wie Abbildung 6 zeigt, ein uneindeutiges Ergebnis.

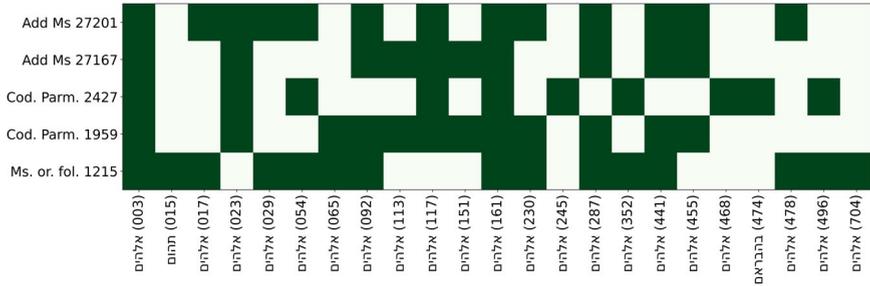
Die etwaigen Dekorationen des Buchstabens *he* mit *Tagin* in den ersten beiden Kapiteln des Toratextes weichen in den erhaltenen Manuskriptzeugen des ‚Sefer Tagin‘ in vielen Fällen voneinander ab. An wenigen Stellen sind sie sich einig, so beim Wort *Elohim* ganz zu Beginn des Toratextes (WortID 003) und in der Mitte des ersten Kapitels (WortID 161). Diese Buchstaben sollen mit jeweils zwei bzw. drei *Tagin* versehen werden – eine Regel, deren Umsetzung in Ms. or. fol. 1215 sichtbar ist. An anderen Stellen (WortID 023 und 117) weisen die ‚Sefer Tagin‘-Varianten eine einheitliche Vorgabe auf, unsere Torarolle weicht jedoch ab.

Die ‚virtuelle Torarolle‘ erlaubt eine weitere quantitative Auswertung, welche die Unterschiede in der Ausführung von Buchstaben und ihren Dekorationen

<sup>18</sup> Vgl. Perani 2022, S. 330 f.; Michaels 2020, S. 5–9.

<sup>19</sup> Der ‚Sefer Tagin‘ liegt in folgenden Manuskriptzeugen vor: Parma, Biblioteca Palatina, Cod. Parm. 1959, Cod. Parm. 2427 und Cod. Parm. 2574; London, British Library, Add MS 27167 und Add MS 27201; Kreuzenstein, Sammlung Graf Wilczek, Ms 5667. Cod. Parm. 2574 und Ms Kreuzenstein 5667 werden im Zuge des Forschungsprojekts kritisch ediert, aber in der vorliegenden ersten Analyse nicht berücksichtigt.

<sup>20</sup> Abkürzung für ‚Text Encoding Initiative‘. Vgl. TEI Consortium 2022.



**Abb. 6 |** Vergleich von Ms. or. fol. 1215 mit ‚Sefer Tagin‘-Textzeugen in Hinblick auf *Tagin* auf dem *he* in Gen 1 und 2. Die y-Achse zeigt die Signaturen der ‚Sefer Tagin‘-Textzeugen und der der Torarolle Ms. or. fol. 1215, die x-Achse Wörter mit *Tagin* (inklusive WortIDs). Markiert wird ein Feld, wenn die Torarolle ein Wort mit *Tagin* auf dem *he* aufweist bzw. der ‚Sefer Tagin‘ ein dekoriertes *he* in einem Wort vorsieht. Wörter ohne *Tagin* in der Torarolle bzw. Wörter, die auch laut ‚Sefer Tagin‘ keine *Tagin* aufweisen sollen, sind für eine übersichtlichere Darstellung ausgelassen. Graphik: ‚ToRoll: Materialisierte Heiligkeit‘.CC BY-NC-ND.

zwischen verschiedenen Torarollen und ‚Sefer Tagin‘-Manuskripten sichtbar macht (Tab. 2). Die Ergebnisse zeigen, dass einige Ausführungen des Buchstabens *he* in den ersten beiden Kapiteln von Genesis in Torarollen zu über 70 oder gar 90 Prozent übereinstimmen, einmal sogar zu 100 Prozent. Die signifikante Übereinstimmung lässt erste Vermutungen zu, dass eine Verbindung zwischen einer bestimmten Manuskripttradition und dem Schreiber einer bestimmten Rolle bestanden haben könnte. Diese Vermutung kann im Rahmen weiterer Analysen von Buchstabenausführungen sowie durch Quellenstudien weiterverfolgt und überprüft werden.

**Tabelle 2 |** Konformität der Dekorationen des *he* in den Kapiteln Gen 1 und 2 von zehn untersuchten Torarollen mit den Dekorationen von vier ‚Sefer Tagin‘-Textzeugen. Angegeben wird absolut und relativ, wie viele der im jeweiligen ‚Sefer Tagin‘ vorgesehenen Dekorationen in den einzelnen Torarollen realisiert wurden.

	Häufigkeit des dekorierten Buchstabens <i>he</i> in Torarollen	Cod. Parm. 1959	Cod. Parm. 2427	Add MS 27167	Add MS 27201
Anzahl zu dekorierender Buchstaben <i>he</i> laut ‚Sefer Tagin‘		12	10	10	13
2° Ms. theol. 1	14	6/12 50%	6/10 60,0%	6/10 60%	10/13 <b>76,9%</b>
Cod. Guelf. 148 Noviss. 2	14	7/12 58,3%	5/10 50,0%	7/10 <b>70%</b>	7/13 53,8%
Cod. Guelf. 149 Noviss. 2	12	8/12 66,7%	6/10 60,0%	7/10 <b>70,0%</b>	9/13 69,2%
Cod. Parm. 3295	11	6/12 50%	6/10 <b>60,0%</b>	5/10 50,0%	7/13 53,8%
Cod. Parm. 3598	9	5/12 41,7%	4/10 40,0%	4/10 40,0%	6/13 <b>46,2%</b>
Cod. Parm. 3599	0	0/12 0,0%	0/10 0,0%	0/10 0,0%	0/13 0,0%
Cod. Parm. 3601	11	9/12 75,0%	5/10 50,0%	7/10 70,0%	10/13 <b>76,9%</b>
Ms. or. fol. 1215	15	7/12 58,3%	5/10 50,0%	5/10 50,0%	10/13 <b>76,9%</b>
Ms. or. fol. 1216	36	11/12 91,7%	9/10 90,0%	9/10 90,0%	13/13 <b>100,0%</b>
Ms. or. fol. 133	11	6/12 50,0%	4/10 40,0%	6/10 60,0%	9/13 <b>69,2%</b>

### 3.5 Analyse der Quellen aus dem Textkorpus der Kommentartradition

Eine für unser Fallbeispiel relevante Quelle ist ein Traktat, der einem einflussreichen Gelehrten des mittelalterlichen Aschkenas zugeschrieben wird. Rabbi Eleazar ben Jehuda aus Worms (ca. 1165 – ca. 1238) gilt als Autor zahlreicher Texte.

Sein literarisches Œuvre umfasst unter anderem religionsgesetzliche, ethische, poetische, esoterische und exegetische Schriften. In einem seiner Werke widmet er sich explizit den *Tagin*. Obgleich auch dieser Text den Titel ‚Sefer Tagin‘ trägt, unterscheidet er sich deutlich von dem älteren oben genannten Schreiberhandbuch gleichen Namens. Eleazars ‚Sefer Tagin‘ ist eine ausführliche Exegese mit zahlreichen Bezügen zu verschiedenen Texten der jüdischen Tradition und im Besonderen zur Heiligen Schrift.<sup>21</sup>

Auf Basis dreier erhaltener Manuskriptzeugen ist auch diese Quelle transkribiert und in eine digitale Edition überführt worden. Für die Edition wurden gemäß den TEI-Richtlinien eigens forschungsrelevante und -spezifische Editionsrichtlinien sowie Auszeichnungsschemata festgelegt. Diese sind flexibel und dynamisch an die Anforderungen der verschiedenen Texte angepasst, dabei zugleich in Hinblick auf die Gesamtedition aller Texte miteinander kompatibel. Je nach Bedarf können die Richtlinien erweitert, verändert und adaptiert werden. Auf diese Weise können die Quellen nicht nur systematisch erfasst und ediert, sondern auch ausgezeichnet und ausgewertet werden. Indem beispielsweise Eigennamen, Orte, Schlüsselbegriffe bzw. Topoi ausgezeichnet werden, erfolgt auch hier zunächst eine quantitative und darauf aufbauend die qualitative Auswertung.

Haben wir zuvor Erkenntnisse über die paläographische Besonderheit des mit drei *Tagin* versehenen Buchstaben *he* mittels einer quantitativen Analyse und einer Auswertung der erfassten Annotationen gewonnen, eröffnet uns die Schrift von Rabbi Eleazar eine weitere Perspektive: Die drei *Tagin* auf dem Buchstaben *he* im Gottesnamen *Elohim* verweisen Eleazar zufolge auf das göttliche Schöpfungswerk, was er mit jeweils drei prägnanten Bibelstellen belegt. Sinngemäß heißt es dort, dass im Anfang drei Dinge geschaffen wurden, nämlich Himmel, Erde und Wasser. Es folgen weitere Ausführungen mit Bezügen zur Dreizahl und sogar gematrische Erklärungen. Ein Aspekt der Heiligen Schrift, in diesem Fall die *Tagin*, wird in einem Zirkelschluss mit dem Bibeltext erklärt und belegt. Für Eleazar birgt der heilige Text der Tora göttliche Geheimnisse und Weisheit, die über den eigentlichen Wortsinn hinausgehen und eben auch in den *Tagin* Ausdruck finden.

Durch Eleazars Ausführungen nähern wir uns dem Phänomen der *Tagin* aus ideengeschichtlicher Sicht. Wenn in einem nächsten Forschungsschritt die aus der Auswertung des Schriftkorpus gewonnenen Erkenntnisse mit denen aus der paläographischen Auswertung der Torarollen abgeglichen und verbunden werden, sind weitergehende Erkenntnisse hinsichtlich der zugrunde liegenden Frage nach dem metaphysischen Potential der *Tagin* und *Otillot Meshunnot* zu erwarten.

---

21 Der Eleazar von Worms zugeschriebene ‚Sefer Tagin‘ ist in bisher drei bekannten Handschriften erhalten: Israel, MS Shunak 3 (Privatbesitz); Oxford, Bodleian Libraries: MS Opp. 540 und MS Opp. 111. Eine kritische Edition des Textes wird im Forschungsprojekt ‚ToRoll: Materialisierte Heiligkeit‘ derzeit vorbereitet.

Voraussichtlich kann so ermittelt werden, ob Eleazar als religiöse Autorität mit seiner Exegese die ausgeübte Schreibpraxis seiner Zeit und Region womöglich beeinflusst oder umgekehrt eine legitimierende Interpretation für eine solche Praxis geboten hat.

#### 4 Konklusion und Ausblick

Unterstützt durch Werkzeuge der *Digital Humanities* zeigt die paläographische Analyse von zehn verschiedenen mittelalterlichen Torarollen einerseits und vier Manuskriptzeugen des ‚Sefer Tagin‘ andererseits die Komplexität des Phänomens der *Tagin* und *Otillot Meshunnot*. Sie bringt zutage, dass viele Schreiber bei der Ausführung des Buchstabens *he* und vieler anderer Buchstaben auf eine Reihe von verschiedenen Schreibtraditionen und Quellen zurückgriffen und diese in ihren Torarollen kreativ umsetzten. Die untersuchten Quellen machen uns mit einer Welt vertraut, in der der Text der fünf Bücher Mose für viele Gelehrte wie Eleazar aus Worms und auch für Schreiber einen hohen Stellenwert als göttliches Gesetz besitzt. Die Wörter und Buchstaben verkörpern das offenbarte Wort Gottes, das sich in und durch die göttliche Schöpfung materialisiert. Die Buchstaben in ihrer Form, ihrem Zahlenwert und ihrer Bedeutung sind als solche heilig und verweisen auf die göttliche Weisheit. Diese ist zugleich in der Heiligen Schrift verborgen. Auf Basis der bisherigen Ergebnisse ist zu vermuten, dass die *Tagin* und *Otillot Meshunnot* sowohl graphischer Bestandteil als auch symbolischer Ausdruck dieser materialisierten Heiligkeit sind. Sie krönen im wahrsten Sinne des Wortes die den heiligen Text konstituierenden Buchstaben.

Textvarianten sowie Inter- und Intratextualität in den vielen verschiedenen Manuskriptzeugen können mittels einer digitalen Edition erfasst und mögliche inhaltliche und rezeptionsgeschichtliche Bezüge sichtbar gemacht werden. Auf diese Weise wird eine Vielzahl von Texten genreübergreifend berücksichtigt und im Verhältnis zueinander ausgewertet. Dieser die Genregrenzen überschreitende Ansatz eignet sich nicht nur für verwandte Themen der Judaistik, sondern auch für Bereiche der Mediävistik, in denen es nicht allein um eine philologische Aufarbeitung bzw. Auswertung von Schriftquellen geht. Die umfangreiche Sammlung der Editionen und Annotationen in einer digitalen Forschungsdateninfrastruktur ermöglicht eine nachhaltige Bereitstellung der Daten und Metadaten sowie eine computergestützte Auswertung und Analyse. Insbesondere die ‚virtuelle Torarolle‘ als zentrale Visualisierungs- und Analyseoberfläche bündelt die Daten und Erkenntnisse der unterschiedlichen Perspektiven und Disziplinen.

Auf Basis der ausgewerteten Forschungsdaten lässt sich bisher konstatieren: *Tagin* und *Otillot Meshunnot* kommen in Wellen. Das bedeutet, dass eine

Regelmäßigkeit bislang nicht erkennbar ist; vielleicht gibt es auch keine. Der Einsatz aller zur Verfügung stehenden digitalen, paläographischen und philologischen Mittel zeigt, dass *Tagin* und *Otijot Meshunnot* nicht nur auf den in der rabbinischen Literatur diskutierten sieben Buchstaben, sondern tatsächlich auf allen 22 Buchstaben und dabei in über 300 verschiedenen Formen vorkommen. Warum es gerade in mittelalterlichen Torarollen zu einem unerlaubten kreativen ‚Regelbruch‘ unter den Schreibern kommt, ist noch nicht abschließend geklärt. Die qualitative Auswertung der Quellen aus dem berücksichtigten Textkorpus deutet jedoch darauf hin, dass insbesondere Buchstaben, Wörter oder Passagen mit besonderer narrativer oder religiöser Bedeutung häufig mit *Tagin* und *Otijot Meshunnot* verziert werden.

Die in diesem Projekt eingesetzten digitalen Werkzeuge basieren auf offenen Standards wie TEI-XML und dem *Web Annotation Data Model*, was eine Nachnutzung und Erweiterung durch andere Forschungsprojekte ermöglicht. Diese Ansätze bieten ein hohes Potential für die transformative Anwendung in vergleichbaren Projekten, etwa in der griechischen oder lateinischen Paläographie. Durch die Veröffentlichung der Methodik und der generierten Daten wird ein Beitrag zur nachhaltigen digitalen Weiterentwicklung geleistet. Die gesamtheitliche Betrachtung eines Phänomens aus unterschiedlichen Perspektiven im Zusammenspiel von klassischen Arbeitsweisen und computergestützten, digitalen Methoden sowie die quantitative und qualitative Auswertung großer disziplinübergreifender Datenmengen eröffnen vielfältige Möglichkeiten und neue Fragestellungen. Dieser Ansatz bietet sich nicht nur für ein spezielles Phänomen wie das der *Tagin* und *Otijot Meshunnot* an, sondern auch für paratextuelle Zeichen oder paläographische Besonderheiten anderer Schriftquellen und Forschungsgebiete.

## Literaturverzeichnis

### Quellen

Berlin, Staatsbibliothek zu Berlin, Preußischer Kulturbesitz: Ms. or. fol. 133.  
<https://resolver.staatsbibliothek-berlin.de/SBB0001D1F100000000>  
 (Zugriff: 14.02.2025).

Berlin, Staatsbibliothek zu Berlin, Preußischer Kulturbesitz: Ms. or. fol. 1215.  
<https://resolver.staatsbibliothek-berlin.de/SBB00005AA700000000> (Zugriff: 14.02.2025)

Berlin, Staatsbibliothek zu Berlin, Preußischer Kulturbesitz: Ms. or. fol. 1216.  
<https://resolver.staatsbibliothek-berlin.de/SBB00005F0A00000000> (Zugriff: 14.02.2025).

**Ganzfried, Shlomo:** Keset ha-Sofer. Ungvar 1871 [ed. princeps 1835].

**Goldschmidt, Lazarus:** Der babylonische Talmud mit Einschluss der vollständigen Mišnah. Berlin, Leipzig 1897–1935.

**Jaakov ben Ascher:** Arbaah Turim. Bd. 1: Oraḥ Ḥajjim. Jerusalem 2018.

**Kagan, Yisrael Meir:** Mishnah Berurah. Jerusalem 2006.

**Karo, Joseph:** Shulḥan Arukh: Oraḥ Ḥajjim. Jerusalem 2018.

Kassel, Universitätsbibliothek Kassel, Landesbibliothek und Murhardsche Bibliothek der Stadt Kassel: 2° Ms. theol. 1.

Kreuzenstein, Sammlung Graf Wilczek: Ms 5667.

**Landsofer, Jonah:** Bnei Jonah. Prag 1803. London, British Library: Add MS 27167, Add MS 27201.

## Forschungsliteratur

**Cruz, António João:** In-Situ Non-Invasive Characterization of Sephardic Torahs from Ponta Delgada, Azores. Poster. TECHNART – International Conference on Analytical Techniques in Art and Cultural Heritage. Lissabon 2023.

**Del Barco, Javier:** From Scroll to Codex: Dynamics of Text Layout Transformation in the Hebrew Bible. In: Bradford A. Anderson (Hg.): From Scrolls to Scrolling Sacred Texts, Materiality, and Dynamic Media Cultures. Berlin 2020, S. 91–118.

**Dunkelgrün, Theodor:** Tabernacles of Text: A Brief Visual History of the Hebrew Bible. In: Ku-ming (Kevin) Chang, Anthony Grafton u. Glenn W. Most (Hgg.): Impagination – Layout and Materiality of Writing and Publication: Interdisciplinary Approaches from East and West. Berlin 2021, S. 47–92.

**Fishbane, Simcha:** The Symbolic Representation of the Sefer Torah. In: Ders.: The Impact of Culture and Cultures Upon Jewish Customs and Rituals: Collected Essays. Boston 2016, S. 211–250.

**Maimonides, Moses:** Mishneh Torah.

Band 2: Sefer Ahavah: Hilkhot Tefillin, Mezuzah ve-Sefer Torah. Jerusalem 2007. Oxford, Bodleian Libraries: MS Opp. 111 und 540.

Parma, Biblioteca Palatina: Cod. Parm. 1959, 2427, 2574, 3295, 3598, 3599 und 3601.

**Shimshon ben Eliezer:** Barukh she-Amar. In: Menachem Mendl Meshi-Zahav (Hg.): Kovetz Sifrei STaM. Jerusalem 1970, S. 15–194.

Wolfenbüttel, Herzog August Bibliothek: Cod. Guelf. 148 Noviss. 2 und Cod. Guelf. 149 Noviss. 2.

**Harris, Steve u. Andy Seaborne:**

SPARQL 1.1 Query Language. W3C Recommendation. 2013. <https://www.w3.org/TR/sparql11-query/> (Zugriff: 08.11.2024).

**Hubmann, Frank u. Josef Oesch:** Betrachtungen zu den Torarollen der Erfurter Handschriften-Sammlung. Untersuchungen zu Gliederung und Sonderzeichen. In: Frank Bussert u. Sarah Laubenstein (Hgg.): Die jüdische Gemeinde von Erfurt und die SchUM-Gemeinden. Kulturelles Erbe und Vernetzung. Jena 2012, S. 96–117.

**Jejkal, Thomas u. Volker Hartmann:** KIT Data Manager. 2020. <https://github.com/kit-data-manager/> (Zugriff: 08.11.2024).

**Kraus, Felix:** EVOKS. 2024. <https://github.com/kit-data-manager/EVOKS> (Zugriff: 08.11.2024).

**Martini, Annett:** „Arbeit des Himmels“. Jüdische Konzeptionen rituellen Schreibens in der europäischen Kultur des Mittelalters. Berlin 2022.

**Michaels, Marc:** Sefer Tagin Fragments from the Cairo Genizah. Leiden 2020.

- Miles, Alistair u. Sean Bechhofer:** SKOS Simple Knowledge Organization System Reference. W3C Recommendation. 2009. <https://www.w3.org/TR/skos-reference/> (Zugriff: 08.11.2024).
- Nehring, Grzegorz, Nehemia Gordon u. Ira Rabin:** Writing and Correcting a Torah Scroll in Germany of the Thirteenth and Fourteenth Centuries. In: *Comparative Oriental Manuscript Studies Bulletin* 7 (2021), S. 123–147. <https://doi.org/10.25592/UHHFDM.9749>.
- Nehring, Grzegorz, Nehemia Gordon u. Ira Rabin:** Distinguishing between Seemingly Identical Inks Using Scanning  $\mu$ XRF and Heat Maps. In: *Journal of Cultural Heritage* 57 (2022), S. 142–148. <https://doi.org/10.1016/j.culher.2022.07.007>.
- Perani, Mauro:** The Tagin: Their Origin, Use, and Oscillating Evolution between Embellishment and Mystical Signifier. New Light from the Ancient Bologna Sefer Torah. In: Élodie Attia u. Antony Perrot (Hgg.): *The Hebrew Bible Manuscripts: A Millennium*. Leiden 2022, S. 297–348.
- Sanderson, Robert:** Web Annotation Protocol. W3C Recommendation. 2017. <https://www.w3.org/TR/2017/REC-annotation-protocol-20170223/> (Zugriff: 08.11.2024).
- Stern, David:** *The Jewish Bible. A Material History*. Seattle 2017.
- TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. 4.4.0. (2022-04-19). <http://www.tei-c.org/Guidelines/P5/> (Zugriff: 08.11.2024).
- Tonne, Danah, Germaine Götzelmann, Nicoletta Pütz u. a.:** TAKITA [Computer software]. 2024. <https://github.com/kit-data-manager/takita> (Zugriff: 08.11.2024).
- Wilkinson, Marc D., Michel Dumontier, IJsbrand J. Aalbersberg u. a.:** The FAIR Guiding Principles for scientific data management and stewardship. In: *Sci Data* 3/160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.
- Young, Benjamin, Paolo Ciccarese u. Robert Sanderson:** Web Annotation Data Model. W3C Recommendation. 2017. <https://www.w3.org/TR/2017/REC-annotation-model-20170223/> (Zugriff: 08.11.2024).

# Annotating Images

## A Prerequisite for Interdisciplinary Research in Medieval Studies

### Contact

#### Dr. Isabella Nicka,

Paris Lodron Universität Salzburg,  
Institut für Realienkunde des  
Mittelalters und der frühen Neuzeit  
(IMAREAL), Körnermarkt 13,  
3500 Krems/Donau, Österreich,  
isabella.nicka@plus.ac.at  
 <https://orcid.org/0000-0002-3587-6606>

#### Miriam Landkammer,

Paris Lodron Universität Salzburg,  
Institut für Realienkunde des  
Mittelalters und der frühen Neuzeit  
(IMAREAL), Körnermarkt 13,  
3500 Krems/Donau, Österreich,  
miriam.landkammer@plus.ac.at  
 <https://orcid.org/0009-0001-3517-0566>

#### Prof. Dr. Andreas Uhl,

Paris Lodron Universität Salzburg,  
Fachbereich Artificial Intelligence and  
Human Interfaces,  
Jakob-Haringer-Str. 2,  
5020 Salzburg, Österreich,  
andreas.uhl@plus.ac.at  
 <https://orcid.org/0000-0002-5921-8755>

#### Michael Linortner,

Paris Lodron Universität Salzburg,  
Fachbereich Artificial Intelligence  
and Human Interfaces,  
Jakob-Haringer-Str. 2,  
5020 Salzburg, Österreich,  
mlinortner@cs.sbg.ac.at  
 <https://orcid.org/0000-0002-6706-4450>

#### Johannes Schuiki,

Paris Lodron Universität Salzburg,  
Fachbereich Artificial Intelligence  
and Human Interfaces,  
Jakob-Haringer-Str. 2,  
5020 Salzburg, Österreich  
jschuiki@cs.sbg.ac.at  
 <https://orcid.org/0000-0002-6197-2327>

**Abstract** Images require an act of interpretation in order to capture their semantic content for (computational) retrieval and analysis, as there is no equivalent process such as transcription for text. The creation of image annotations is, therefore, a complex undertaking but a prerequisite not only for the application of digital methods to art historical research but also for the inclusion of images in cross-media and cross-disciplinary analyses in medieval studies. In this article, we will first outline the main issues and challenges of annotating medieval images. Based on the results of a study on depicted wood in medieval art, we will then focus on relevant processes and requirements in the annotation of visual phenomena and their evaluation. We will also explore the differences in perception of painted wood textures between humans and computer vision algorithms on the one hand, and project participants and laypersons on the other.

**Keywords** Art History; Computer Vision; Digital Medieval Studies; Image Annotation; Distant Viewing

## 1 What Significance Does Image Annotation Have for Digital Humanities Approaches in Medieval Studies?

Many studies in art history are based on qualitative, in-depth analyses of a relatively small number of works of art. This is likely to change as digitisation campaigns become more ambitious and comprehensive, as the pool of available visual and textual resources about the past continues to grow, and as

more cases emerge that demonstrate how annotation can lead to new research questions. One of the main benefits of such a development is that creating processable annotation data on medieval artworks may (finally) become a focus of art history and result in increased collaboration with computer science to achieve the goal of making more annotation data available for interdisciplinary research.

For now, however, digital approaches to art history can build on available artwork metadata (artist, date, technique, location, etc.) or on what computer vision calls ‘low-level feature’ image information. The latter is what can be derived directly from a digital photograph of a work of art (colour saturation, etc.). The epistemological potential of analysing these data in a humanities context is rather limited.<sup>1</sup>

The situation is different in the text domain: because text is composed of a certain number of alphanumeric characters and a limited number of structural elements (paragraphs, line breaks, headers, etc.), it is possible to create computable transcripts, although the effort required to produce them may vary. Many computational approaches can then start from analysis of the words in the transcripts themselves (word frequency and their distribution in the text, topic modelling) or from automatic annotation by comparing tokens (natural language processing, named entity recognition, etc.).<sup>2</sup> As noted by Arnold and Tilton, “there is no equivalent option, as there is with structured and textual data, to delay interpretation of the images’ meaning(s) before applying computational methods.”<sup>3</sup> In other words, when it comes to comparing certain semantic aspects of images and integrating insights at this level into interdisciplinary research in the field, image annotation and how it is modelled is key to applying Digital Humanities (DH) approaches to medieval studies.<sup>4</sup>

In this paper, the terms ‘metadata’ and ‘annotation’ are used to describe object metadata, which is information about an entire work of art, and local annotation, which is additional information added to one aspect – for instance – a single depicted element in a pictorial unit.<sup>5</sup>

Once image annotation data is available, it can be added to the pool of other (partly pre-existing) machine-readable data on texts, archival records, musical notations, archaeological finds, etc., through which comparative studies can be carried out. There may be instances where such an analysis already reveals an existing trans-medial discourse, making it possible to explore it across media

---

1 For an example of such analyses, see Pippich 2015.

2 Cf. Schubert 2023.

3 Arnold and Tilton 2023, p. 23.

4 For brief insights on image annotations in the Digital Humanities, see Rapp 2017, pp. 264f.; Lordick, Becker, Bender et al. 2016, pp. 190f.

5 Cf. Schöch 2017, pp. 228–229.

boundaries. In other cases, it will be necessary due to the different media conditions to uncover strategies, patterns, or particularities in the treatment of certain topics within one medial form and then to compare the findings with strategies in other media which are part of the discourse. Each situation can be found in the following examples:<sup>6</sup>

The retable on the high altar in Pöggstall (Lower Austria) shows a figure in the background of Christ's flagellation, handing a coin to another character. If the annotation 'coin' is available for this image and it is linked to the concept of money, it can be compared with the flagellation in the text of a particular religious play (the so-called St. Gall Passion Play). Here, Rufus promises a reward of twenty marks for a particularly cruel flogger. If we were to add the concept of 'money' to the term 'marks', we would be able to trace a common denominator for the two medial instances, suggesting an overlap of discourse.

An example of a specific strategy in images is a retable that otherwise operates by setting the Passion events in the Lower Austria of the time ('Vergegenwärtigung') but which uses a pseudo-Hebraic cross titulus to explicitly indicate Jerusalem as the place of Christ's burial. If we can identify this strategy through adequate annotation, we can then compare it with the strategies in the textual sources and see how/if there are also ways of switching between the events of the Passion in the historical location of Jerusalem and the contemporary world of the faithful.

In this article, we will first outline the main issues and challenges of annotating medieval images and then focus on the annotation of visual phenomena using a case study of painted depictions of the wood of Christ's cross. Next, we will discuss relevant processes in the manual creation of such annotations by different actors and the application of Artificial Intelligence (AI) algorithms, and show how quantitative analyses of the annotations can be used in art history and as a prerequisite for interdisciplinary research using Digital Humanities methods in medieval studies.

## 2 Annotating Medieval Images

Image annotation can cover different areas, such as depicted entities and their relations, and visual phenomena – or, to put it more succinctly, what is represented in figurative images and how.<sup>7</sup>

---

6 For more details on these examples, see Nicka 2014, pp. 280–282; Nicka 2019, pp. 110–119.

7 Another relevant area of image annotation in art history would be, e.g. interdependencies between images (transfer of motifs, spatial organisation, etc.).

When annotating depicted entities in images, the need quickly arises to integrate them into a more complex data model in order to represent core visual statements and use them in interdisciplinary research.

- As mentioned above, the first step in image annotation on a semantic level is an act of interpretation: one needs to make sense of the part of the picture one wants to annotate, e.g. flower, book cover, St. Barbara. These examples show that linking a part of an image to a verbal concept can be done in different ways, depending, among other things, on the specificity of the concept, from very general terms to distinct individuals or objects, or even categories for depicted surface textures of materials (see Sections 3–4). Moreover, one must consider whether the annotation refers to a whole entity or to a part of an entity (e.g. a book clasp as part of a book) as well as how much contextual knowledge<sup>8</sup> is needed (e.g. to identify St. Barbara).
- In texts, the word ‘flower’, for example, does not necessarily need to be specified. In contrast, one of the medial conditions of images is that even a rudimentary sketch must outline some details of a flower (e.g. flower petals) so that it can be recognised. This makes it difficult to decide at what level of granularity to annotate.
- Unlike linear text, the order in which entities are perceived in an image is not fixed.<sup>9</sup> For instance, a connection between a person on the left of the image throwing an apple to a person standing on the far right cannot be inferred from their respective coordinates in the pictorial space but would have to be coded for further analysis.
- As for roles in actions, e.g. when one person throws an apple to another person, the respective active or passive involvement is not made explicit by annotating the depicted elements with ‘person A’, ‘person B’, ‘apple’, and ‘throwing’, whereas in texts – at least to some extent – a kind of differentiation is provided by grammatical structures.

While the broader pursuit of image annotation modelling, ontologies for image content,<sup>10</sup> and existing domain specific classification systems (e.g. ICONCLASS) or authority files<sup>11</sup> is beyond the scope of this paper, we would like to refer briefly to two projects with specific data models for image annotations.

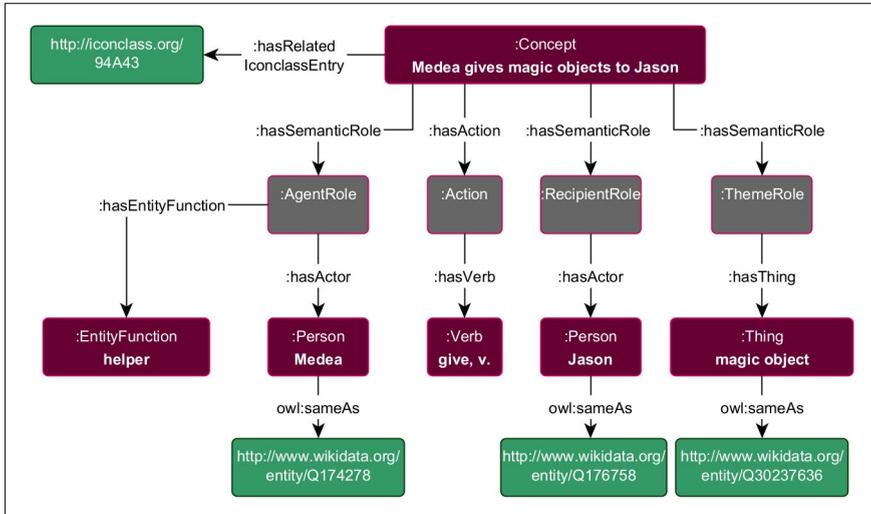
---

<sup>8</sup> Cf. Noll 2016, pp. 194f., for a description of Panofsky’s three levels of reading an image (pre-iconographic/iconographic/iconologic).

<sup>9</sup> Cf. Hinkelmanns, Landkammer, Nicka et al. 2021, pp. 157–160.

<sup>10</sup> E.g. Gartner 2020 or Sartini, Baroncini, van Erp et al. 2023.

<sup>11</sup> Cf. Kailus and Stein 2018, pp. 132–134.



**Fig. 1 |** A narrative from the story of the Argonauts as recorded in the ‘Ontology of Narratives of the Middle Ages’ (ONAMA).

In ‘REALonline’,<sup>12</sup> an image database of medieval and early modern visual media from Austria and its adjacent regions in neighbouring countries, a scene graph is used to capture depicted entities, their parts, and some forms of relationships between entities (e.g. a jug standing on a table). There are also nodes for information about the setting (e.g. interior), actions, or places (e.g. Salzburg). Further annotations in ‘REALonline’ cover specific properties of depicted image elements for which, in computer vision, the term ‘object qualification’ would apply (as complementary to object detection and object localisation).<sup>13</sup> For characters, additional information is added to record the individual’s name, gender, status, occupation, gestures, and postures. Colour, material, and shape are given for non-person entities. Links to authority files and classification systems (like GND or ICONCLASS) allow us to infer further information about the depicted entities. With approximately 22,500 records and more than 1.2 million image annotations on a semantic level, it is possible to retrieve images with a number of different parameters as well as their distribution in the ‘REALonline’ corpus (e.g. the query above, ‘find and count images of the flagellation of Christ with a person holding a coin’).

<sup>12</sup> realonline.imareal.sbg.ac.at (accessed: 05/11/2024). For further information, see Nicka 2019; imareal.sbg.ac.at/imareal-projekte/realonline (accessed: 05/11/2024).

<sup>13</sup> Arnold and Tilton 2023, p. 39.

The second project is ONAMA ('Ontology of Narratives of the Middle Ages').<sup>14</sup> In this interdisciplinary collaboration between German literature studies and art history, a Semantic Web ontology was created to record narrative building blocks of images and texts. Based on a concept in linguistics, semantic roles were assigned to the actors, objects, places, or temporal elements depending on an action, so that, for example, an actor (agent) could be distinguished from the person affected by an action (patient) (Fig. 1). Narrative functions that entities have only in a specific situation or sequence (e.g. gift, opponent, etc.) can also be recorded in ONAMA. This makes it possible to describe and analyze – across media boundaries – the roles of individual narrative elements in the story and to ask questions such as 'find narratives in images or texts in which an animal (role: agent) brings something to someone'.

### 3 Annotating Visual Phenomena I: Vocabulary, Workflow, Insights

In the remainder of this paper, we will consider how visual phenomena in medieval paintings, such as wood texture patterns, can be annotated. This is work that has hardly been touched upon and will allow us to better understand the functions of these visual markings. To do so, we will draw on our experiences in the project KIKI ('How Material Came into the Picture: Exploring Cultural Innovations Inter-disciplinarily with AI and DH methods'), which focuses on the visual appearance of depicted materials – in particular, wood.<sup>15</sup>

Wood is a natural, biogenic material, and unlike some other materials (e.g. gold, brick, snow), its appearance is highly individual due to the variety of the tree, its age when cut down, the impacts of nature and humans during the tree's life, and the processing (e.g. cutting direction and degree of surface finishing). In a painted rendition, the range of visual appearances of wooden surfaces is further increased due to the artistic license of the artist.

In the KIKI project, computational methods are explored in order to investigate the use of wood in pictorial strategies on a broader basis. We manually annotated an image corpus, delineating the depictions of wooden surfaces. These polygon annotations form the ground truth for the texture classification experiments with computer vision methods (see Section 4). The annotations are furthermore evaluated in order to analyse the representations of wood in the corpus: by applying

---

<sup>14</sup> [onama.sbg.ac.at](http://onama.sbg.ac.at) (accessed: 05/11/2024). For further information, see Hinkelmanns, Landkammer, Nicka et al. 2021.

<sup>15</sup> [imareal.sbg.ac.at/en/imareal-projects/how-material-came-into-the-picture-kiki/](http://imareal.sbg.ac.at/en/imareal-projects/how-material-came-into-the-picture-kiki/) (accessed: 05/11/2024). KIKI was funded by the Federal State of Salzburg and is a collaboration between art history and computer vision.

a distant-viewing approach, visualisations of the annotation data, combined with existing metadata, can be used to detect possible patterns, trends, and outliers. The term ‘distant viewing’ was first used by BENDER for a quantitative approach to art history, analogous to ‘distant reading’ in literary history.<sup>16</sup> In our project, we use it to refer to an exploratory data analysis of manually generated annotation data and metadata on images through visualisations of statistics, which can be useful even in a small data set of depicted materials in medieval paintings in order to understand the data and uncover areas of inquiry.<sup>17</sup> Following that, artworks can be analysed with a close-viewing approach in order to explore reasons and backgrounds.

We also examined the value of vocabulary and polygon annotations for art-historical material-/wood-related analyses. In the following, we discuss the development of the vocabulary and the annotation method and provide a brief insight into the analysis.

### 3.1 Finding Words for Patterns

For our comparisons, the nuances of wood grain patterns demand more specific classes than simply ‘wood’. This raises the question of how to label painted surface textures of an optically heterogenous material with verbal categories. Classification according to tree species would only be possible, if at all, in the case of paintings done in a highly naturalistic style. We therefore looked at how wood grain patterns – more correctly: figures – are described in the field of woodworking. One way wood grain patterns differ is in how the log is cut: if it is cut tangentially, the growth rings will appear as a ‘cathedral’ grain pattern. If it is cut radially through the centre, the growth rings will emerge as a straight grain pattern (cf. Fig. 2). The categories ‘cathedral’ and ‘straight’ can be adopted for depicted wood grains. Obviously, in depicted timber, all grain patterns can be represented on any face of a beam or board. Painters made use of this leeway, which is why we only use terms that indicate a pattern and not, for example, ‘tangential face’ or ‘end grain’. Due to the frequency of the represented ‘cathedral’ figure, we have made further subdivisions not used in woodworking terminology: rounded, pointed, or balanced appearance. We have also found ‘burl’ to be a usable class for very distinct irregular wood grains. However, there are no concepts used in woodworking or carpentry that are applicable to the various dotted or dashed grain patterns

---

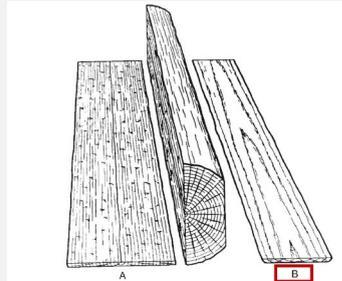
<sup>16</sup> Bender 2015.

<sup>17</sup> On distant viewing as a method drawing on the concept of exploratory data analysis (EDA), see Arnold and Tilton 2023, esp. pp. 50–52. On the potential of data visualisations to reveal paths of investigation by generating “productive uncertainty” even in small data sets, as they often occur in medieval studies, see Schwandt and Wachter 2024.

### Textur „gefladert“

Wenn eine Objektfläche parabelförmige Kurven aufweist, als „gefladert(-spitze Formen/runde Formen/ausgewogen)“ annotieren.

Hintergrund: Als Fladerung bezeichnet man eine Holzzeichnung, bei der die Jahresringe als parabelförmige Kurven auf der Holzoberfläche erscheinen. Sie zeigt sich, wenn das Holz im Tangentialschnitt (Längsschnitt außerhalb der Stammmitte) geschnitten wird.



Public domain, [https://commons.wikimedia.org/wiki/File:Plain\\_quarter\\_sawn.png](https://commons.wikimedia.org/wiki/File:Plain_quarter_sawn.png). Source: Forest products laboratory (ed.): Wood handbook – wood as an engineering material. Madison WI 2021, p. 3-15.

**Fig. 2 |** Page on the ‘cathedral’ wood grain pattern from the KIKI project’s annotation guideline. The diagram shows boards cut from a log: quartersawn (a, showing a straight grain pattern) and plainsawn (b, showing ‘cathedral’ grain).



**Fig. 3 |** Examples for the category ‘dotted/dashed/tiger stripe’ from the annotation guideline for depicted wood grain patterns. REALonline 005319, 001845, 000448. IMAREAL – University of Salzburg.

depicted in medieval paintings. These patterns result from different anatomical factors in wood (e.g. cut-through medullary rays of some types of wood or different structural anomalies). These can most often not be distinguished in the case of painted renditions, which is why we have introduced a class combining general terms describing a pattern and specialist terminology (‘dotted/dashed/tiger stripe’, Fig. 3). Since experiential knowledge of a material, on the one hand,

and representational conventions, narrative functions, ideas about the material, compositional reasons, etc., on the other, may contribute to the appearance of a material in a painting,<sup>18</sup> a combination of material processing terminology and general pattern description terms seems all the more suitable. An expert from the field of wood technology, Michael GRABNER (Vienna), kindly agreed to clarify some technical questions.

How similar two paintings of wood grains look depends on more than just the figure; we went for an additional classification according to the dimensions of ‘regularity’, ‘contrast’, ‘degree of abstraction’, and ‘sophistication of texture rendering’.<sup>19</sup> Additionally, we have identified two other factors in all represented types of wood grain that have an impact on the visual appearance of a depicted wooden surface: the existence of differing heartwood colour and of cracks.

### 3.2 Workflows for Annotating Visual Markings

We formed a corpus of images from the 14th and 15th centuries from the ‘REAL-online’ database which show the cross of Christ in the context of different subjects or other wooden crosses in religious narratives. To ensure consistency in the annotations, we drafted an annotation guideline defining what qualifies as a depiction of wood within the project. The scope seems obvious, but a definition is necessary: for instance, it is possible that an object is depicted entirely without a surface texture or even with an abstract texture, but in a colour that leads us to perceive it as wood. Conversely, a depicted texture reminiscent of wood might coincide with an unlikely colour (see, for an example, the greenish wood grain on one cross in Fig. 8), or even if neither the represented texture nor the colour correspond to actual wood, we still understand it to be wood based on our knowledge of the depicted object. We therefore concluded that a texture reminiscent of wood grain on at least part of the cross must be present.

The guideline also describes the texture classes and the other attributes using images and brief descriptions (Fig. 2). Bark was included as a texture category,

---

<sup>18</sup> Cf. Nicka 2023.

<sup>19</sup> In psychological experiments that investigate – partly in relation to the development of computational texture classification methods – which high-level dimensions (i.e. abstract attributes that go beyond the specific textures of certain materials) are decisive in the rapid evaluation of the similarity of textures, certain constants emerge despite varying results. For instance, Cho, Yang and Hallett 2000 considered a perceptual texture space of four or five dimensions, where the six attributes – contrast, coarseness, directionality, regularity, lightness, and edginess – are useful directions in that space. Rao and Lohse 1996 postulated three dimensions that are significant to texture perception, namely repetitiveness, contrast/directionality, and granularity/coarseness/complexity.

owing to the cross being the object investigated, but was not further subdivided. In addition to the category descriptions, it was necessary to specify how the objects were to be outlined for the computer vision methods (single shapes per each plane on the object, which are then grouped together; handling of ‘interferences’ such as drops of blood on the wood).

From the corpus of 502 images, 287 met the criteria (wood grain depicted, sufficient photo quality or state of preservation). Within these, 409 crosses were annotated, consisting of a total of 4,106 individual shapes (Fig. 4).<sup>20</sup> After aggregating the data (categorical tags, pixel areas), it was merged with metadata from the ‘REALonline’ database.

### 3.3 Insights from Annotations of Visual Appearances

The bar chart on occurrence and image area of wood grains (Fig. 5) shows that segmentation is essential in a ‘distant viewing’ of depicted materials, as there are large discrepancies between the two parameters: most strikingly, ‘no texture’ is the most frequent texture, while the ‘cathedral-balanced’ texture has by far the largest area within the total of all images. Wood grain is often missing on the foreshortened planes that are numerous but occupy small pixel areas, such as the underside of the horizontal beam. We evaluated the wood grain of each largest annotated shape – usually on the primary visible face of the cross and of significance for the visual appearance – of the cross of Christ and, as a particular variant of it, the True Cross (the relic of the cross found by Empress Helena in Jerusalem 300 years after Christ’s death, which in turn was later seized by the Persian ruler and regained by Heraclius).<sup>21</sup> The analysis also includes other crosses shown in scenes of the Passion of Christ (subsumed under ‘cross thief’) and the crosses of St. Andrew, St. Peter, and St. Philip (Fig. 6).

The stacked bar chart indicates that ‘cathedral’ grain (and here predominantly ‘balanced’) occurs most frequently on all distinguished crosses. Furthermore, it becomes apparent that a significant proportion of the thieves’ crosses shows bark (27 of 78 crosses = 34.6 percent), while bark is very rarely found on the cross of Christ (7 out of 255 crosses = 2.7 percent). This distribution discloses that the suggested effort put into making the crosses is a common pictorial strategy to emphasise the cross of Christ or Christ himself in images where several crosses are shown (Fig. 4 gives an example). Having identified this distribution, further exploration can be undertaken concerning the narrative contexts where this

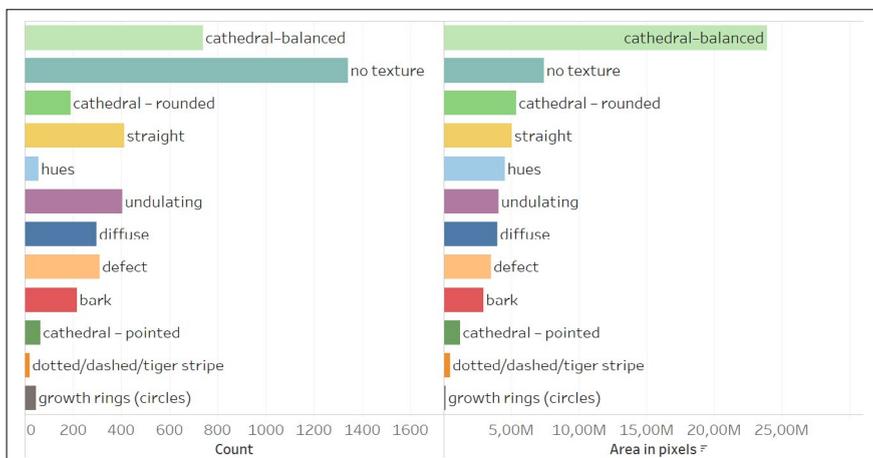
---

<sup>20</sup> Annotation was carried out using the Computer Vision Annotation Tool ([www.cvat.ai/](http://www.cvat.ai/) [accessed: 05/11/2024]).

<sup>21</sup> Cf. Toussaint 2011, pp. 53–59.

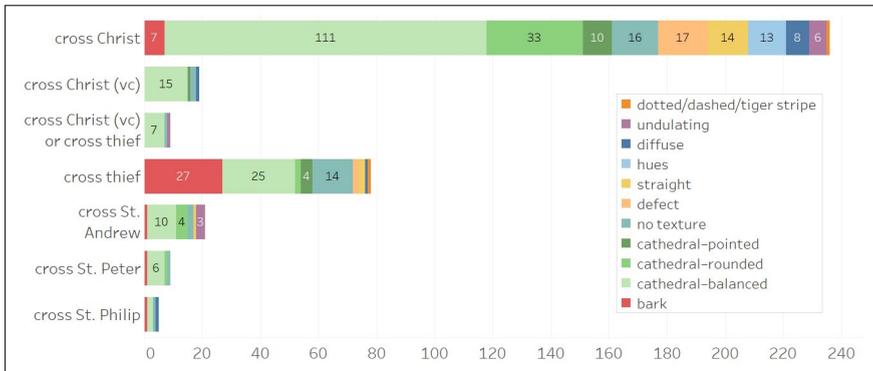


**Fig. 4 |** Polygon annotation of depicted wood in the Computer Vision Annotation Tool (CVAT). Image credits for medieval image: IMAREAL – University of Salzburg.



**Fig. 5 |** Distribution of surface textures in 287 annotated images showing crosses, according to the number of shapes (left) compared to the area in pixels covered (right). The category ‘burl’ is missing: it was not assigned once. ‘Defect’ refers to imperfections in the painted surface.

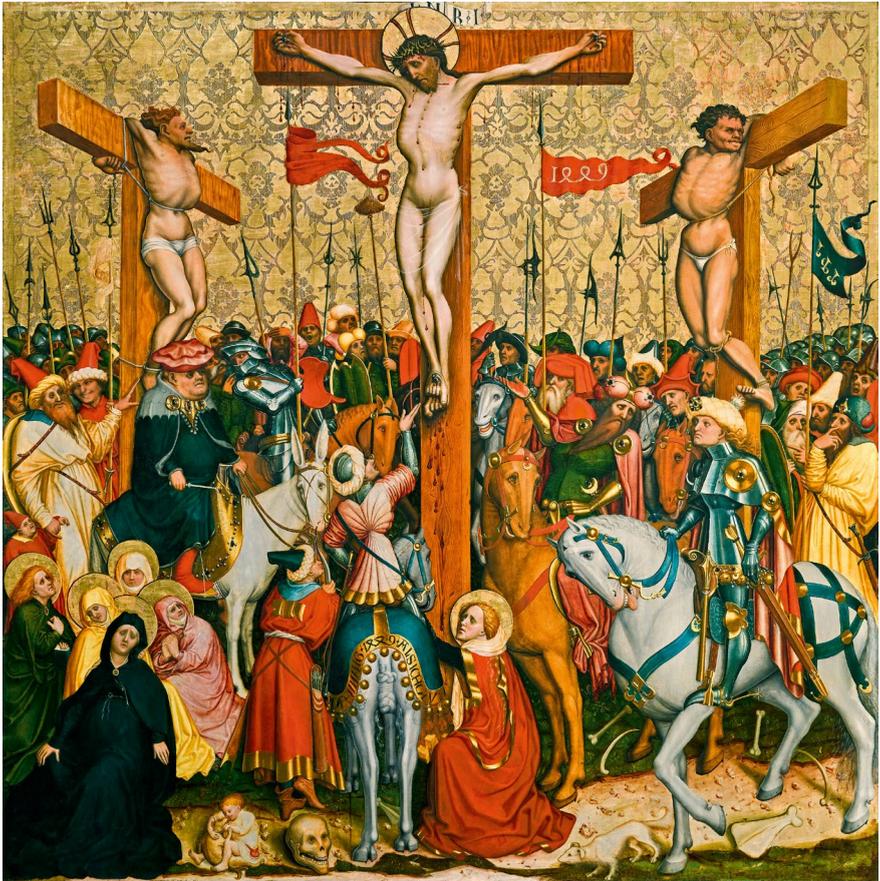
differentiation occurs: it is found in images of the crucifixion of Christ, in scenes preceding and following the crucifixion, and also when Helena finds three crosses underground. Of particular interest as an outlier is the texture ‘dotted/dashed/tiger stripe’, with the lowest number in total. It is only found very occasionally on the cross of Christ and on the cross of one thief: within those paintings, we find a striking use of material representation as a narrative strategy in ‘Crucifixion of Christ’ by Conrad Laib (Fig. 7, Vienna, Belvedere, 1449). In the qualitative analysis of the artwork, we have identified a distinctive pattern in the horizontal beam of



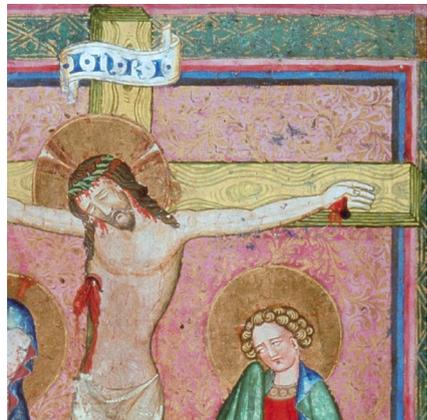
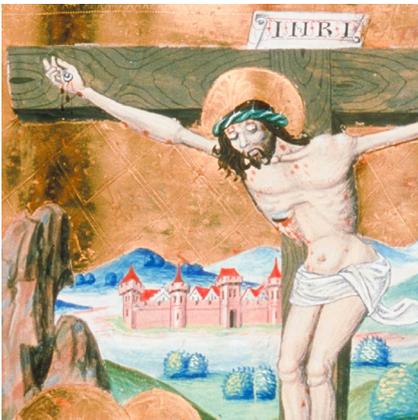
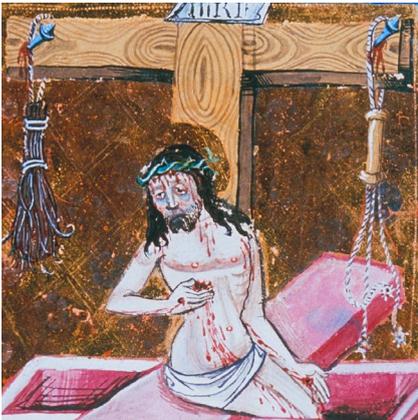
**Fig. 6 |** Distribution of surface textures of annotated crosses (only the largest annotated shape per cross) among the persons from the religious narratives.

Christ's cross, which echoes the blood running down from Christ's wounds and also reinforces the link with the good thief, whose cross-beam has the same grain. Together, they form a contrast to the cross of the impenitent thief. Textures of the class 'dotted/dashed/tiger stripe' are found far more frequently in wooden furniture surfaces depicted in paintings from the 14th and 15th centuries, as seen in images from the 'REALonline' database. Hardwood may have been an inspiration, as opposed to softwood where the growth rings predominate in the figure. The same applies to 'burl', which is sometimes found on depicted furniture surfaces in 'REALonline' but which the data analysis reveals does not occur in the annotated corpus of wooden crosses.

As shown in this section, categorical annotation allows for the identification of broad trends in the representation of wooden surfaces. It is only partially suitable for comparing wood grains in detail (such as the various manifestations of the most frequently occurring 'cathedral' grain). Figure 8 shows five crosses where surfaces on the front side were annotated with the same combination of attributes ('cathedral - rounded', 'sophistication of the painted texture rendering: coarser', 'degree of abstraction: higher', 'regularity: strong', 'contrast: strong'). While the first two examples (in reading direction from top left) actually look very similar and even stem from the same codex, the third example also shows clear similarities but differs in colour (we did not annotate colour values), and the last two examples deviate significantly both from the first three examples and from each other.



**Fig. 7 |** Top: Conrad Laib, 'Crucifixion of Christ', 1449, probably from the former parish church (now Franciscan church) in Salzburg. Belvedere Vienna. CC BY-NC-SA 4.0. Bottom: details of the Laib Crucifixion. Cf. REALonline 000718.



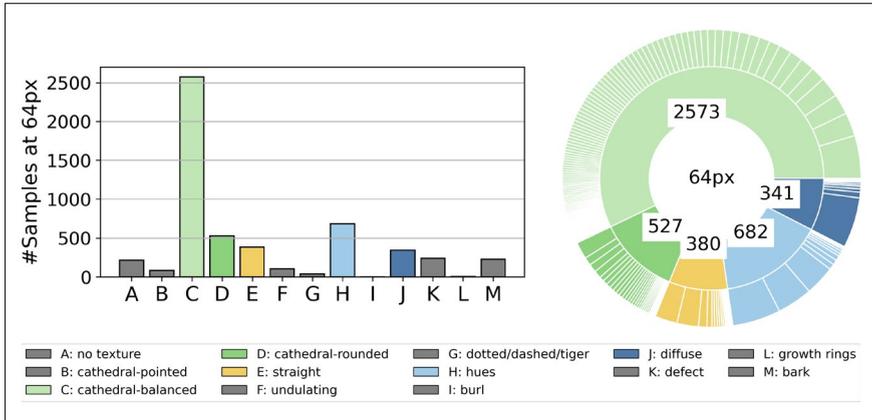
**Fig. 8** | Details from selected images where surfaces on the front side of the cross were annotated with the same texture category and further attributes. REALonline 006049, 006064, 005755, 007820, 009855. IMAREAL - University of Salzburg.

## 4 Annotating Visual Phenomena II: Agents and Annotation Quality

Who creates image annotations? There are three answers to this question: experts and collaborators in a (research) project, laypeople (with little or no preliminary schooling for annotating in a citizen science application, or performing on-demand tasks as paid crowdworkers), and computer algorithms. In the previous section of this paper, we have already discussed the role of experts and annotators within a research project and how they build the vocabulary, enforce guidelines for a consistent annotation process and shape the output. Based on experiments in the KIKI project, in the following section we will shed light on the differences between human and algorithmic production of annotations, and between lay people in a citizen science scenario and experts/collaborators in a research project.

### 4.1 Human | Computer Algorithm

This section looks at how the perception of painted wooden texture coincides or differs between humans and computer vision algorithms. To do so, patches are cropped from the manually annotated areas showing depicting wooden structures in 287 images using various patch sizes, such as  $64 \times 64$  pixels or  $128 \times 128$  pixels. Dividing the data into smaller patches is a common strategy for texture classification because it artificially increases the dataset, while it is assumed that the defining structures are fine-grained and can thus be found on every patch. Furthermore, the computer vision methods used in the following part operate on quadratic input images, making quadratic patches a natural choice (as opposed to squeezing images into the right resolution, possibly destroying relevant structures during the process). Figure 9 shows the number of patches extracted from the annotations with a patch side length of 64 pixels. Due to unbalanced data and very few patches being available for some classes, only five texture classes (indicated by coloured bars) are used for further analysis. The nested pie chart on the right visualises the distribution of cropped patches based on their origin image, where each segment in the outer ring of a texture class represents the proportion of samples originating from the same source image. The number of patches per source image varies both because of differing resolutions of the images and differently sized wooden crosses within the paintings. The data distribution presents a challenge for learning-based methods in further experiments: first, the imbalance of the textures has to be kept in mind when choosing a strategy for splitting data into training and test samples; second, patches stemming from the same source image should never be in both training and test sets, as neighbouring patches could be too similar and therefore distort the outcome of the experiments. However, for some texture classes, most of the patches originate from just a few source images (e.g. Classes H and J).



**Fig. 9** | Left: number of samples per texture class. Coloured bars: five classes used for further evaluation. Grey bars: remaining texture classes. Right: distribution of texture types (inner circle) divided into chunks indicating the number of samples cropped from the same source image (outer circle).

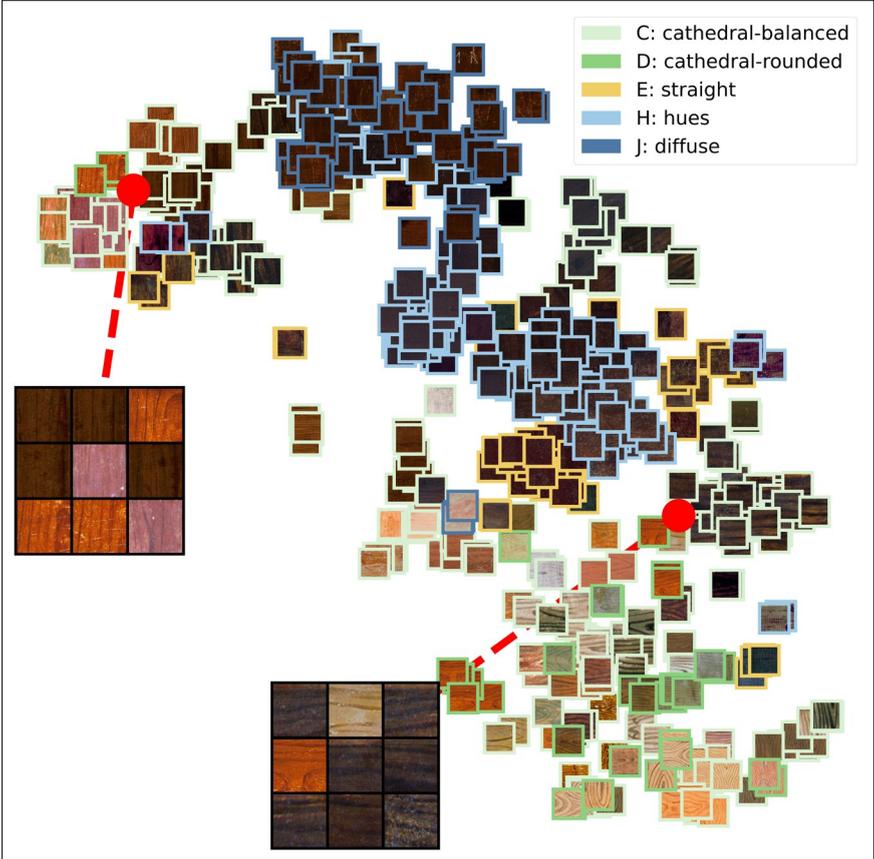
To see how well the computer algorithm's classification of patches matches the human annotation (ground truth), state-of-the-art methods from the domain of computer vision are utilised: for every patch, a 2,048-dimensional feature vector is extracted by employing a ResNet-50<sup>22</sup> (pre-trained on the ImageNet image database)<sup>23</sup> where the last layer (classification layer) is omitted. In this way, each image patch is encoded in a 2,048-dimensional feature vector. Next, t-SNE<sup>24</sup> is used to reduce the dimensionality to a two-dimensional space. The resulting 2D vectors are then used as coordinates and the corresponding patches visualised in figure 10.

Images closer in 2D space seem more related to the algorithm. The coloured border for each patch indicates the ground truth obtained via human annotation. Clusters of patches with similar annotation labels are visible, but it is important to consider that this could be because they stem from the same neighbourhood in the source image and, thus, are almost identical. It is, however, interesting to see that distinct structures in the texture patches, such as horizontal or vertical lines, are placed predominantly in neighbouring regions in the 2D space (see the zoomed example patches). While, for many applications, out-of-the-box deep learning features deliver satisfactory results, the t-SNE visualisation hints that the features of ResNet-50 fail to sufficiently embed the relevant structures in our texture classification setup.

<sup>22</sup> Cf. He, Zhang, Ren et al. 2015.

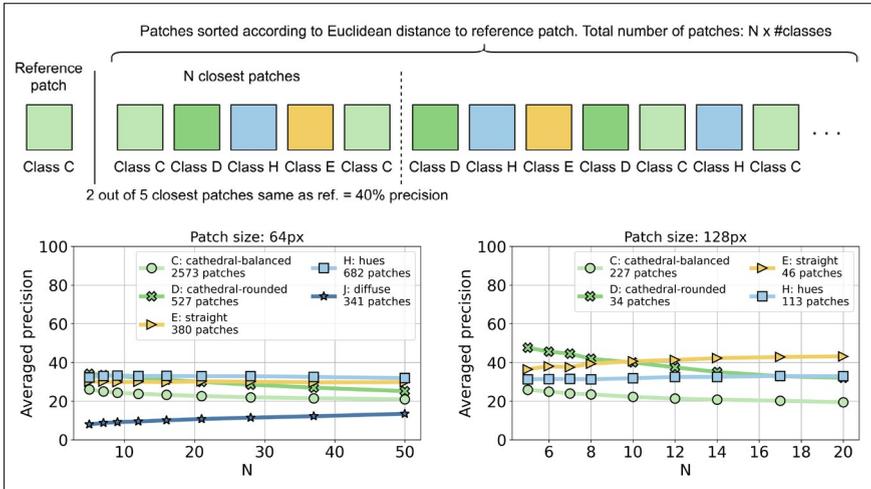
<sup>23</sup> Cf. Deng, Dong, Socher et al. 2009.

<sup>24</sup> Cf. van der Maaten and Hinton 2008.



**Fig. 10** | t-SNE visualization of ResNet-50 feature embeddings using 128 px samples. Borders are coloured according to their annotated texture class.

For further analysis, image retrieval experiments are carried out: for every patch,  $N$  other patches are randomly selected from every texture class, while patches from the same source image are omitted. This ensures a balanced and unbiased experimental setup. The selected patches are then sorted according to the Euclidean distance between their ResNet-50 embedding and the one embedding of the reference patch. Subsequently, the proportion of the  $N$  closest patches in the same texture category as the reference patch (also called precision) is used as an evaluation metric. The bottom plots in figure 11 show the precision values averaged over all patches. For the patches with 128 pixels side length, class J is omitted because forty-three of the forty-six available patches belong to the same source image. Because four and five classes are used for the retrieval experiments, 25 percent and 20 percent precision, respectively, would be random guessing.



**Fig. 11** | Top: visual explanation of the retrieval experiment. Bottom: retrieval experiment with patch sizes 64 px and 128 px, respectively.

In contrast to many other domains, out-of-the-box deep learning methodology does not seem to work properly here. We identify two possible reasons for unsatisfactory results:

- (i) As a consequence of patching, it is possible that structures relevant for the classification of the painted wood texture get destroyed.
- (ii) Texture classification tasks in computer vision usually work on fine-grained, repeating patterns. In comparison, the structures for the texture of painted wood are rather big.

## 4.2 Citizen Science | Science

To evaluate the human perception of texture samples from painted wood, a small-scale experiment was carried out including non-experts. In a science-to-public event, fifty-two participants were given three reference examples per texture class and were tasked to assign ten randomly picked samples to a texture type. A screenshot from the setup can be seen on the left in figure 12. The right plot in figure 12 shows the ground truth texture classes on the left and the texture classes assigned by the participants on the right, wherein coinciding texture classes are shown using more saturated colours. The numbers on the left indicate the total

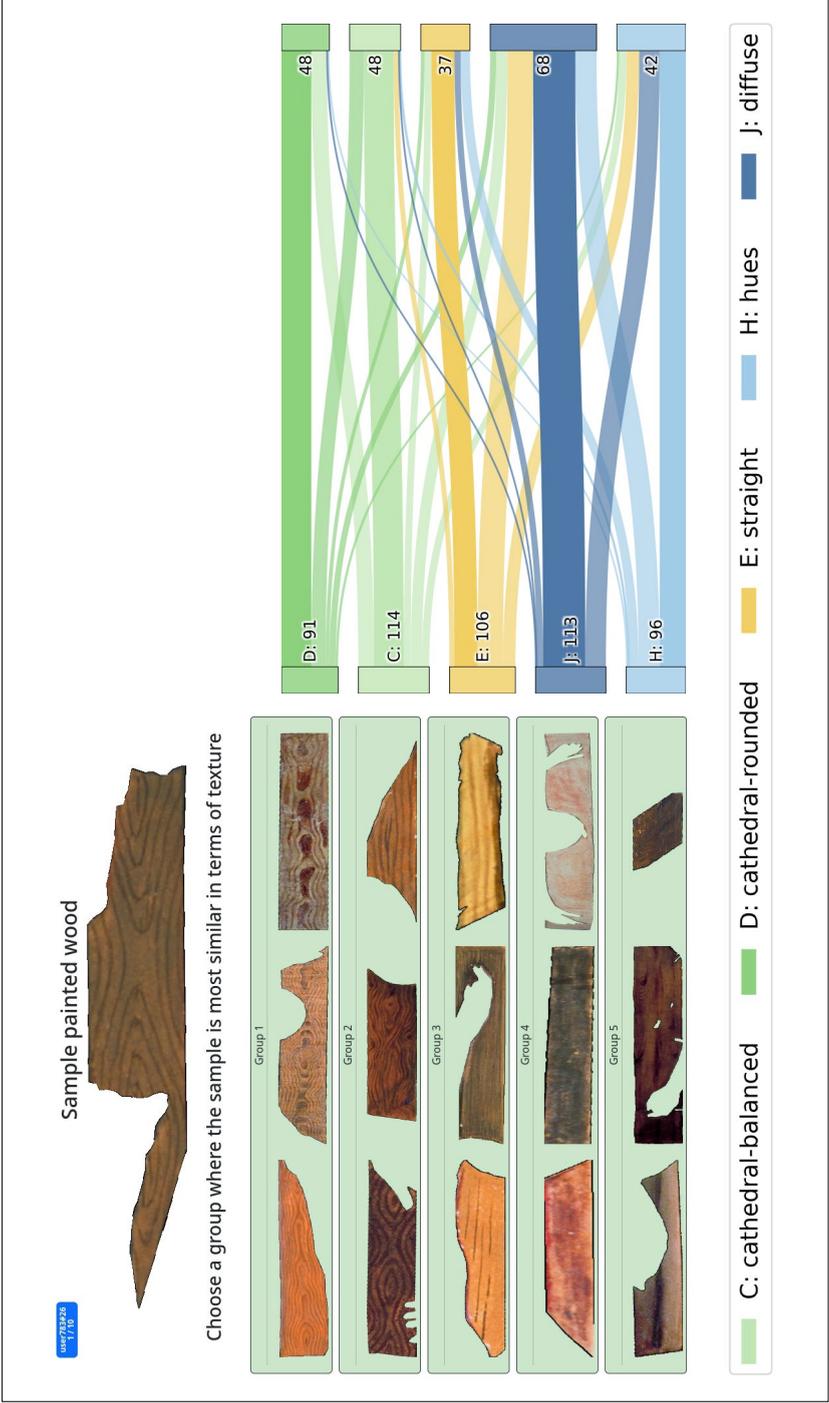


Fig. 12 | t-SNE visualization of ResNet-50 feature embeddings using 128 px samples. Borders are coloured according to their annotated texture class.

occurrence of the texture classes throughout the experiment, and the numbers on the right display the number of samples where the assigned class and the ground truth class coincide. Samples were drawn randomly from a uniform distribution of the classes, i.e. they did not follow the distribution shown in figure 9. The diagram provides a visual representation of how well the participants matched the samples with the ground truth labels. About half of the samples were assigned according to the expert annotation. This shows that textures were perceived differently between participants, but it is still far from random guessing, which would be 20 percent consensus. When looking at the biggest flow to a non-matching category, one can deduce that some categories were perceived to be similar, for example categories D and C.

## 5 Closing Remarks

As has been shown, image annotation can be used to demonstrate the recurrence or exceptionality of pictorial strategies: in his panel painting of the Crucifixion from Salzburg (Fig. 7), Conrad Laib goes his own way in his choice of the same texture for the crosses of Christ and of the good thief, as opposed to the differing texture of the cross of the unrepentant thief. Only when we have annotations at our disposal to disclose such narrative strategies that can only be realised by means specific to images – the differentiation of crosses by the degree of their surface finishing was found to occur frequently – will comparisons with strategies used in texts, by means inherent to them, become possible. For example, we could apply an apt computational analysis method and explore how the wood of Christ's cross is emphasised in late medieval Passion treatises, where other properties, such as the scent of the wood, play a major role.<sup>25</sup> From our experience so far with 'REALonline' and ONAMA, a more complex modelling of the annotations will be useful to capture more elaborate visual strategies. Of course, this could also be approached in a multi-stage manner, using computational methods to process more basic image annotations. As the comparison of expert annotations and computer algorithms has shown, further efforts will be needed to adapt standard AI methods for annotating visual phenomena in historical images. With respect to citizen science projects, the findings suggest that less ambiguous annotation tasks may be more opportune for generating consistent data. While some processes in image annotation are the same as in text annotation, in particular the creation of the vocabulary used – the need to apply words to visual phenomena – is more difficult, and the results yielded, when trying to

---

<sup>25</sup> Cf. Kemper 2006.

differentiate between fine-grained visual structures, are not always fully satisfactory. In contrast to texts, not only is it more laborious to address the sections to be annotated but the segmentation is important for the ground truth in the computer vision part. Annotation data on exact image regions, as this paper has shown, allow for a sounder basis when it comes to analysing visual phenomena like depicted materials.

## Bibliography

- Arnold, Taylor and Lauren Tilton:** Distant Viewing: Computational Exploration of Digital Images. Cambridge MA 2023. <https://doi.org/10.7551/mitpress/14046.001.0001> (accessed: 07/06/2024).
- Bender, K.:** Distant Viewing in Art History: A Case Study of Artistic Productivity. In: *International Journal for Digital Art History* 1 (2015). <https://doi.org/10.11588/dah.2015.1.21639> (accessed: 07/06/2024).
- Cho, Raymond Y., Vicky Yang and Peter E. Hallett:** Reliability and Dimensionality of Judgments of Visually Textured Materials. In: *Perception & Psychophysics* 62 (2000), pp. 735–752.
- Deng, Jia, Wei Dong, Richard Socher et al.:** Imagenet: A Large-Scale Hierarchical Image Database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255.
- Gartner, Richard:** Towards an Ontology-Based Iconography. In: *Digital Scholarship in the Humanities* 35 (2020), pp. 43–53. <https://doi.org/10.1093/llc/fqz009> (accessed: 08/11/2024).
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren et al.:** Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778.
- Hinkelmans, Peter, Miriam Landkammer, Isabella Nicka et al.:** Beyond the Plot. Der Vergleich mittelalterlicher Narrative im Semantic Web mit ONAMA. In: *Vienna Doctoral Academy – “Medieval Academy”* (ed.): *Narrare – producere – ordinare. Neue Zugänge zum Mittelalter* (Wiener philologisch-kulturwissenschaftliche Studien 1). Wien 2021, pp. 145–160. <https://doi.org/10.23783/9783706910637> (accessed 28/10/2024).
- Kailus, Angela and Regine Stein:** Besser vernetzt. Über den Mehrwert von Standards und Normdaten zur Bilderschließung: Informationszugang, semantische Interoperabilität, Linked Open Data, Normdaten, Kореferenzierung. In: *Piotr Kuroczyński, Peter Bell and Lisa Dieckmann* (eds.): *Computing Art Reader. Einführung in die Digitale Kunstgeschichte*. (Computing in Art and Architecture 1). Heidelberg 2018, pp. 118–139. <https://doi.org/10.11588/arthistoricum.413.c5772> (accessed 18/10/2024).
- Kemper, Tobias A.:** Die Kreuzigung Christi. Motivgeschichtliche Studien zu lateinischen und deutschen Passionstraktaten des Spätmittelalters (Münchener Texte und Untersuchungen zur deutschen Literatur des Mittelalters 131). Tübingen 2006.
- Lordick, Harald, Rainer Becker, Michael Bender et al.:** Digitale Annotationen

- in der geisteswissenschaftlichen Praxis. In: *Bibliothek Forschung und Praxis* 40 (2016), pp. 186–199. <https://doi.org/10.1515/bfp-2016-0042> (accessed 07/06/2024).
- Maaten, Laurens van der and Geoffrey Hinton:** Visualizing Data Using t-SNE. In: *Journal of Machine Learning Research* 9 (2008), pp. 2579–2605.
- Nicka, Isabella:** Interfaces. Berührungszonen von Transzendenz und Immanenz im spätmittelalterlichen Sakralraum. In: Marion Meyer and Deborah Klimburg-Salter (eds.): *Visualisierungen von Kult.* Wien, Köln, Weimar 2014, pp. 260–293. Figures: pp. 438–444. <https://doi.org/10.7767/boehlau.9783205793052.260> (accessed 07/06/2024).
- Nicka, Isabella:** Object Links in/zu Bildern mit REALonline analysieren. In: Institut für Realienkunde des Mittelalters und der frühen Neuzeit (ed.): *Object Links – Dinge in Beziehung* (formate – Forschungen zur Materiellen Kultur 1). Wien 2019, pp. 95–126. <https://doi.org/10.7767/9783205209591.95> (accessed 07/06/2024).
- Nicka, Isabella:** Stein und Erde gestalten. Die *Concordantiae Caritatis* als Experimentierfeld eines Distant Viewings zur Materialdarstellung im 14. Jahrhundert. In: *MEMO* 10 (2023): Material aspektivieren, pp. 120–160. <https://doi.org/10.25536/20231005> (accessed: 07/06/2024).
- Noll, Thomas:** Ikonographie/Ikonologie. In: Ulrich Pfisterer (ed.): *Metzler Lexikon Kunstwissenschaft. Ideen, Methoden, Begriffe.* Stuttgart 2016, pp. 194–198. <https://doi.org/10.1007/978-3-476-00331-7> (accessed: 18/10/2024).
- Pippich, Waltraud von:** Rot rechnen. In: Constanze Baum and Thomas Stäcker (eds.): *Grenzen und Möglichkeiten der Digital Humanities* (Sonderband der Zeitschrift für digitale Geisteswissenschaften 1). 2015. [https://www.doi.org/10.17175/sb001\\_016](https://www.doi.org/10.17175/sb001_016) (accessed: 07/06/2024).
- Rao, A. Ravishankar and Gerald L. Lohse:** Towards a Texture Naming System: Identifying Relevant Dimensions of Texture. In: *Vision Research* 36 (1996), pp. 1649–1669.
- Rapp, Andrea:** Manuelle und automatische Annotation. In: Fotis Jannidis, Hubertus Kohle and Malte Rehbein (eds.): *Digital Humanities. Eine Einführung.* Stuttgart 2017, pp. 253–267. [https://www.doi.org/10.1007/978-3-476-05446-3\\_18](https://www.doi.org/10.1007/978-3-476-05446-3_18) (accessed: 07/06/2024).
- Sartini, Bruno and Sofia Baroncini, Marieke van Erp et al.:** ICON: An Ontology for Comprehensive Artistic Interpretations. In: *Journal on Computing and Cultural Heritage* 16/3 (2023). <https://doi.org/10.1145/3594724> (accessed: 17/03/2025).
- Schöch, Christof:** Aufbau von Datensammlungen. In: Fotis Jannidis, Hubertus Kohle and Malte Rehbein (eds.): *Digital Humanities. Eine Einführung.* Stuttgart 2017, pp. 223–232. [https://doi.org/10.1007/978-3-476-05446-3\\_16](https://doi.org/10.1007/978-3-476-05446-3_16) (accessed: 07/06/2024).
- Schubert, Charlotte:** Textanalyse digital. In: Christina Antenhofer, Christoph Kühberger and Arno Strohmeyer (eds.): *Digital Humanities in den Geschichtswissenschaften.* Wien 2023, pp. 144–165. <https://elibrary.utb.de/doi/10.36198/9783838561165-144-165> (accessed: 07/06/2024).
- Schwandt, Silke and Christian Wachter:** Visualization as Irritation: Producing Knowledge About Medieval Courts Through Uncertainty. In: *Frontiers in Big*

Data 7 (2024). <https://doi.org/10.3389/fdata.2024.1188620> (accessed: 07/06/2024).

**Toussaint, Gia:** Kreuz und Knochen.  
Reliquien zur Zeit der Kreuzzüge.

Berlin 2011. <https://doi.org/10.3389/fdata.2024.1188620> (accessed: 07/06/2024).

# ITEM

## Eine Schnittstelle realienkundlicher Daten aus historischen Textquellen

### Kontakt

**PD Dr. Elisabeth Gruber,**  
Paris Lodron Universität Salzburg,  
Institut für Realienkunde des  
Mittelalters und der frühen Neuzeit,  
Körnermarkt 13,  
3500 Krems/Donau, Österreich,  
elisabeth.gruber2@plus.ac.at  
 <https://orcid.org/0000-0003-1286-8256>

**Peter Färberböck,**  
Paris Lodron Universität Salzburg,  
Institut für Realienkunde des  
Mittelalters und der frühen Neuzeit,  
Körnermarkt 13,  
3500 Krems/Donau, Österreich,  
peter.faeerberboeck@plus.ac.at  
 <https://orcid.org/0000-0002-8701-8259>

**Abstract** At the Institute for Medieval and Early Modern Material Culture (IMAREAL), the ‘ITEM’ research tool is designed to be an interface that makes two types of written sources—inventories and account books—accessible for comprehensive material culture queries. This tool enables users to explore and query objects and their attributes, such as materials, values, or contexts of use, from these two types of sources. Both types are systematically annotated and supplemented with metadata to facilitate their integration into Linked Open Data (LOD) and, in the long term, to contribute to a shared semantic web of material culture data.

Due to the heterogeneity of these sources, thesauri and ontology are ‘mapped’ with the flexibility and fundamental concept of extending this mapping to other source types as well. Here, LOD is understood not merely as a connection to controlled vocabularies but as a means of linking diverse data sets and projects within medieval studies, thus enabling a systematic exploration of various source types concerning material culture inquiries.

**Keywords** Linked Open Data; Material Culture; Digital Medieval Studies; Inventories; Bookkeeping

## 1 Objekte in Texten pragmatischer Schriftlichkeit: Inventare und Rechnungsbücher

Die Auseinandersetzung mit Gegenständen „im Schnittpunkt sozialer Beziehungsnetze“ hat seit der Pionierarbeit von Katharina SIMON-MUSCHEID unterschiedliche Schwerpunktsetzungen und Ausformungen erfahren.<sup>1</sup> Zuletzt verwies Christina ANTENHOFER, ausgehend von den deutsch-italienischen fürstlichen Eheverbindungen des 14. und 15. Jahrhunderts, ausführlich auf das Beziehungen stiftende Potential materieller Kultur des Mittelalters und der Renaissance.<sup>2</sup> Dorothee RIPPMMANN hat in einem Beitrag zu Leben, Arbeit und materieller Kultur im Kontext pragmatischer Schriftlichkeit an Schweizer Beispielen erneut deutlich gemacht, dass auch Rechnungsbücher als Texte zu verstehen seien und demnach zum Umgang mit materieller Kultur befragt werden können.<sup>3</sup>

Die Auswertung von Inventaren im Hinblick auf die Frage, ob und wie materielle Objekte Bedeutung stiften, steht am Institut für Realienkunde des Mittelalters und der frühen Neuzeit (IMAREAL) schon längere Zeit im Mittelpunkt des Interesses. Insbesondere das Verhältnis zwischen der Benennung von Räumen, den Objekten, die als in den Räumen befindlich beschrieben wurden, und möglichen Nutzungsszenarien von Räumen und Objekten wurde stichprobenartig untersucht.<sup>4</sup> Als Quellengattung eignen sich raumbezogene Inventare vor allem für qualitative Auswertungen von Objekt-Konstellationen, die in einem spezifischen räumlich-zeitlichen Setting aufgezeichnet wurden. So werden etwa zum Zeitpunkt des Ablebens, anlässlich der Übernahme eines Erbes, anlässlich von Herrschaftsübergaben oder zu anderen, meist mehr oder weniger ausführlich definierten Gelegenheiten Aufzeichnungen angefertigt, die den zu übergebenden mobilen Besitz beschreiben. Im Zentrum des Interesses steht dabei der Dokumentationscharakter dieser spezifischen Ansammlung von mobilen Gegenständen zu einem bestimmten Zeitpunkt. Das Ensemble an Gegenständen wird in Form einer Momentaufnahme fixiert und verschriftlicht. Variabel ist dabei die Ausführlichkeit der anlassbezogenen Beschreibung der Objekte hinsichtlich ihrer materiellen und/oder ideellen Beschaffenheit und Wertigkeit. Demgegenüber haben Rechnungsbücher die Aufgabe, die Finanzverwaltung der rechnungsführenden

---

1 Simon-Muscheid 2004.

2 Vgl. Antenhofer 2021.

3 Vgl. Rippmann 2015, S. 210; zuletzt: Gruber 2015 sowie Gruber 2019.

4 So wurden beispielsweise im Projekt ‚RaumOrdnungen‘ Gegenstände aus schriftlichen Quellen und archäologischen Fundkontexten erschlossen und in Beziehung zu Räumen, Personen und Handlungen gesetzt: <https://www.imareal.sbg.ac.at/imareal-projekte/raumordnungen/> (Zugriff: 20.03.2025). Im Detail dazu: Handzel u. Kühnreiter 2015; Handzel, Schichta u. Schmid 2015; Kühnreiter 2019; Matschinegg 2019.

Einheit zu verschriftlichen. Getätigte Ausgaben werden den verbuchten Einnahmen gegenübergestellt und die monetären Werte mit den zugehörigen materiellen und immateriellen Leistungen beschrieben. Beide Quellengattungen – Inventare wie Rechnungsbücher – bieten Informationen zu Gegenständen, sind jedoch in unterschiedlicher Weise mit Zusatzinformationen wie beispielsweise Wertigkeiten, Beschaffenheit, räumlicher Position oder personeller Zuordnung angereichert.

Der bereits digital erschlossene Bestand an Burginventaren bildet aufgrund der Dichte an verzeichneten Objekten eine wichtige Grundlage für Aussagen über Ausstattungsmuster von Burgen und für die Analyse von Objektgesellschaften.<sup>5</sup> Über das Erfassungstool ‚ITEM/inventories‘<sup>6</sup> sind diese Inventardaten nun in Form einer graphbasierten Datenbank zugänglich. Aktuell sind 31 Inventare mit ca. 800 Räumen und 12.000 Einzelobjekten sowie Angaben zu deren Bewertung, Stückzahl, Materialbeschaffenheit und weiteren Eigenschaften der verzeichneten Gegenstände abfragbar. Diese listenförmig angeführten Objekte ermöglichen aufgrund der Dichte Aussagen zu Ausstattungsmustern von Burgen, da nur Inventare dieses Gebäudetyps erfasst wurden.

Materialien, verarbeitete Rohstoffe und Gegenstände lassen sich als Warengruppen auch aus Rechnungsbüchern erschließen. Die Verschriftlichung von Einnahmen- und Ausgabenverzeichnissen setzte mehrheitlich im Spätmittelalter ein und wurde im Laufe der Zeit kontinuierlich systematisiert.<sup>7</sup> Neben Gesamtrechnungen, die unterschiedlichste Aspekte der Rechnungsführung abbilden, sind vor allem in Sonderrechnungen häufiger spezifische Cluster von materiellen Objekten zusammen mit ihrem Warenwert verzeichnet, da sie auf einen konkreten Anlass bezogen sind. So bieten beispielsweise Baurechnungen die Möglichkeit, ein spezifisches Ensemble an Gegenständen, Rohstoffen und Dienstleistungen für die Durchführung eines Bauprojektes auszuwerten. Sie ergänzen damit die auf die Ausstattung von Burgen bezogenen Inventarien hinsichtlich des Begriffsfeldes ‚bauen‘ – ‚wohnen‘ – ‚arbeiten‘, das für einige am IMAREAL bearbeitete Projekte von Interesse ist.<sup>8</sup> Für die realienkundliche Erschließung von Rechnungsbüchern wurde daher ein weiterer Workflow entwickelt, der es ermöglicht, objektbezogene

---

5 Ein instruktives Fallbeispiel bietet Matschinegg 2019, die Objekte und Orte des Schreibens auf Burgen analysiert.

6 Ein Launch von ‚ITEM‘ ist im zweiten Quartal 2025 geplant: <https://item.imareal.sbg.ac.at> (Zugriff: 20.03.2025).

7 Für einen Überblick vgl. Gleba u. Peterson 2015; Gruber 2015; Vogeler 2015; Vogeler 2021.

8 Zu nennen ist hier neben dem Projekt ‚Der domestizierte Raum‘ (<https://www.imareal.sbg.ac.at/imareal-projekte/der-domestizierte-raum/>) auch das Projekt ‚Klosterhöfe in der Wachau‘ (<https://www.imareal.sbg.ac.at/imareal-projekte/klosterhoeft-in-der-wachau/>) (Zugriff: 24.04.2025), das sich mit den materiellen Aspekten klösterlicher Wirtschaftshöfe in der Wachau beschäftigt; vgl. Gruber u. Kührtreiber 2021.

Daten aus Rechnungsbüchern über das Online-Tool ‚ITEM/accounts‘ darzustellen und abfragbar zu machen.

## 2 Realienkundliche Daten in Texten pragmatischer Schriftlichkeit

Neben den vorgestellten Tools ‚ITEM/inventories‘ und ‚ITEM/accounts‘ entsteht nun unter dem Arbeitstitel ‚ITEM‘ ein online zugängliches Abfrage-Tool für Inventare und Rechnungsbücher. Es ermöglicht eine gemeinsame digital unterstützte Erfassung und Analyse von Gegenständen und Materialien beider dem Bereich der pragmatischen Schriftlichkeit zugeordneten Quellengattungen.<sup>9</sup> Der Name des Tools verweist dabei auf zwei Aspekte, die die Zielsetzung des Projekts veranschaulichen: Zum einen kennzeichnet der sowohl in Inventaren als auch in Rechnungsbüchern verwendete lateinische Begriff *item* (‚ebenso‘) die Zugehörigkeit eines Eintrags zu einem spezifischen Geschäftsfall, das könnte im Fall des Inventars beispielsweise die Erfassung aller Gegenstände in einem Raum oder im Fall des Rechnungsbuches einer Gruppe von getätigten Ausgaben sein.<sup>10</sup> Zum anderen beschreibt der englischsprachige Begriff nicht nur das *object of interest*, sondern hat mit seinem Bedeutungsumfang Eingang in die Fachsprache des digitalen Arbeitens gefunden. Diese semantischen Anspielungen greift der gemeinsame Name des Tools ‚ITEM‘ auf, welches die Daten als *Linked Open Data* (LOD) zu Objekten verbindet, die aus zwei verschiedenen Quellengattungen mit Aufzählungsstruktur stammen, und Forschungsstränge bedient, die auf ähnlich strukturierte Daten zugreifen.

Mit dem 2019 initiierten Relaunch der Projektdatenbank ‚RaumOrdnungen‘ (2007–2012), bei der die in ausgewählten Inventaren aufgelisteten Objekte raumbezogen und strukturiert erfasst wurden, liegt eine graphbasierte Abfragemöglichkeit der aufgenommenen Daten vor.<sup>11</sup> Die in den Inventaren listenförmig angeführten Objekte ermöglichen aufgrund der Dichte Aussagen zu Mustern der Ausstattung von – in diesem ersten Fall – Burgen. In Kooperation mit dem Projekt

9 Für die kulturwissenschaftliche Forschung ist die begriffliche Unterscheidung von Gegenständen und Objekten von Bedeutung. Insbesondere Matschinegg 2019, S. 76, weist auf die vielfältigen Perspektiven hin, die durch theoriegeleitete Zugänge ermöglicht werden.

10 Auf die inhaltliche Bandbreite der Bedeutung des Begriffs im Kontext von Rechnungsbüchern verweisen ausführlich Gleba u. Peterson 2015, S. 7–11; zum Umgang mit der Quellengruppe Rechnungen vgl. Gruber 2015; Vogeler 2015.

11 Zum Projekt ‚RaumOrdnungen‘ vgl. Handzel, Schichta u. Schmid 2015; Handzel u. Kühnreiter 2015; Handzel, Klug u. Schmid 2010 sowie Kühnreiter 2005. Der Relaunch wurde maßgeblich von Ingrid Matschinegg geleitet und unter Mitwirkung von Doris Hörmann, Elias Knapp und Tobias Pamer durchgeführt. Ermöglicht wurde er durch die finanzielle Unterstützung der Abteilung Wissenschaft und Kultur des Landes Niederösterreich, die stets großes Interesse an den Projekten des IMAREAL im Bereich der *Digital Humanities* zeigt.

„INVENTARIA: Inventarisierung als soziale Praxis“ wurde der bereits bestehende „RaumOrdnungen“-Thesaurus verfeinert und erweitert.<sup>12</sup> Dieser erfasst nun in drei Stufen von der groben Raumkategorie bis zur genauen (durch die Quelle überlieferten) Bezeichnung des Raumes die begrifflichen Varianten der Raumbezeichnungen. Die Kooperation mit dem digitalen Kulturerbe-Projekt „Hohensalzburg digital“, das die schriftliche Überlieferung der materiellen Raumausstattung der Festung Hohensalzburg mit aus einem 3D-Modell rekonstruierten Gebäudedaten zusammenführt, arbeitet ebenfalls mit der begrifflichen Grundstruktur von „RaumOrdnungen“.<sup>13</sup> Da „ITEM“ von Grund auf als Anlaufstelle für *Digital Humanities*-Projekte zu (Burg-)Inventaren und in Zukunft auch zu Rechnungsbüchern konzipiert ist, wurde dieser Thesaurus mit diesem Hintergrund erstellt und kommt nun auch zur Anwendung.

Als Pilotprojekt zur Erweiterung der Datengrundlage um Rechnungsbücher dient ein bereits ediertes Rechnungsbuch des 15. Jahrhunderts, das mehrere Jahre andauernde Baumaßnahmen abbildet. Mit Hilfe der für unterschiedliche Arten von Rechnungsbüchern entwickelten „Bookkeeping-Ontology“<sup>14</sup> wird der für Inventare erstellte Thesaurus mit dem Rechnungsbuch-Thesaurus gemappt. Durch die Normdaten-Verknüpfungen entsteht ein Datensatz, der diese Quellengattungen gemeinsam abfragbar macht und somit eine Grundlage für weitere Fragen und Antworten aus Inventaren und Rechnungsbüchern bietet. Dieser Ansatz erlaubt erstmals, systematisch Fragen zu Objekten an diese Quellen (gleichzeitig) zu stellen und sie gemeinsam zu durchsuchen – offen für alle und verbunden via RDF.<sup>15</sup>

Das Forschungsinteresse am IMAREAL zielt darauf ab, herauszufinden, wie die in den gewählten Fallbeispielen generierten Daten hinsichtlich der Fragen nach dem sozialen Raum modelliert und wie sie ausgewertet werden können. Die Grundidee, „ITEM“ als Schnittstelle zu konzipieren, die zwei Typen schriftlicher Quellen – Inventare und Rechnungsbücher – enthält, ermöglicht es, Objekte und deren Eigenschaften aus diesen beiden Perspektiven über beide Quellengattungen

---

12 Zum Projekt „INVENTARIA“ vgl. Antenhofer, Gruber-Tokić, Hiebel u. a. 2023; Antenhofer u. Matschinegg 2021 sowie Gruber-Tokić, Hiebel, Danthine u. a. 2023.

13 Für Informationen zum Projekt „Hohensalzburg digital“ vgl. Antenhofer, Matschinegg, Brandstätter u. a. 2023.

14 Vgl. Pollin 2019. Zur Ontologie vgl. Eide, Pollin u. Vogeler 2015. Die „Bookkeeping-Ontology“ wurde am Institut für Digitale Geisteswissenschaften (ehemals Zentrum für Informationsmodellierung) von der *Digital Edition Publishing Cooperative for Historical Accounts* (DEPCHA) in Graz entwickelt.

15 RDF oder *Resource Description Framework* ist ein Datenmodell zur Darstellung und für den Austausch von Daten im Web und damit eine grundlegende Beschreibungsvorschrift für das *Semantic Web*. In diesem Framework werden Aussagen in Tripel-Form (Subjekt, Prädikat und Objekt) gespeichert, um sie strukturiert abfragen zu können. Bei „ITEM“ bieten wir einen Export in dieses Format an, arbeiten aber in der Datenbank selbst mit der Open-Source-Graphdatenbank Neo4j. Vgl. dazu: Pollin 2021.

hinweg abzufragen. Beide Quellengattungen werden entsprechend systematisch annotiert und mit Metadaten versehen, damit sie für LOD und in weiterer Folge für ein gemeinsames *Semantic Web* realienkundlicher Daten genutzt werden können. Wie solche übergreifenden Abfragen auf der Basis von Thesaurusbegriffen historischer Quellen neue Ansätze ermöglichen, wird am Beispiel der Analyse von Holz als Material deutlich.

### 3 Use Case: Holz als Material

Im Rahmen der Forschungsperspektive ‚Materialities‘ wird am Institut für Realienkunde des Mittelalters und der frühen Neuzeit die materielle Kultur über Materialien als Werkstoffe und deren Grundlage untersucht. Besonderes Augenmerk wird dabei auf die Beziehungen zwischen Menschen und Materialien gelegt, die multiperspektivisch betrachtet werden.<sup>16</sup> Wir gehen davon aus, dass die Eigenschaften der Materialien erst im konkreten Funktions- und Bedeutungszusammenhang fassbar werden und dass sie erst durch die Verwendungszusammenhänge Bedeutung gewinnen und diese gegebenenfalls auch verändern. Sie sind „Konstituenten von Welterfahrung und Weltbildern und somit grundlegende Bausteine von historischen Prozessen“<sup>17</sup>. Gerade in der Vormoderne ist Holz Ausgangsmaterial für zahlreiche Gegenstände verschiedenster Gebrauchskontexte. Geschirr, Waffen- und Jagdgerät, Spielzeug und Haushaltsgeräte, Möbel und Werkzeuge basieren auf diesem Grundstoff, der in unterschiedlicher Weise bearbeitet wurde.<sup>18</sup> Als organisches Material unterliegt Holz jedoch potentiell einem raschen Zerfallsprozess, sei es durch Verwitterung oder Entsorgung als Brennmaterial. Alltagsgegenstände aus Holz wie etwa Teller, Schüsseln oder Handgriffe von Werkzeugen und Waffen sind daher nur in geringer Anzahl als Objekte überliefert. Häufiger sind diese in schriftlichen Aufzeichnungen zu finden, etwa wenn ihr Erwerb in Rechnungsbüchern verzeichnet ist oder ihr Bestand in Inventaren dokumentiert wird. In diesen Überlieferungszusammenhängen werden die Gegenstände an einem bestimmten Abschnitt ihres Lebenszyklus fassbar. Ausgangspunkt einer Untersuchung könnte beispielsweise eine qualitative Abfrage

---

<sup>16</sup> Vgl. dazu das Themenheft von MEMO 1 (2017): Holz in der Vormoderne. Werk-Stoff, Wirk-Stoff, Kunst-Stoff, <http://dx.doi.org/10.25536/2523-2932012017>; einen aktualisierten Forschungsüberblick aus der Perspektive der *Material Culture Studies* bieten Kühnreiter u. Schlie 2017; allgemein zu Holz als ‚Naturstoff‘ vgl. Radkau 2007, bes. S. 7–130.

<sup>17</sup> Kühnreiter u. Schlie 2017, S. 4.

<sup>18</sup> Vgl. Wunschel 2018.

sein, die beide Quellengattungen nach Gegenständen durchsucht, bei denen als Ausgangsmaterial ‚Holz‘ eingetragen ist.<sup>19</sup>

Eine einfache Facetten-Suchabfrage<sup>20</sup> zum Thesaurusbegriff ‚Holz‘ im Recherchetool ‚ITEM/inventories‘ ergibt 152 Ergebnisse, die sich auf Gegenstände beziehen, die als aus Holz bestehend bezeichnet werden oder deren Holzcharakter aufgrund einer validen Interpretation erschlossen wurde. So wird beispielsweise im Inventar der Burg Wurmberg (1525) eine aus dem Holz einer Zypresse gefertigte SchlieÙe (Tassel) – ein *zipresseins tasl* – aufgelistet, die gemeinsam mit einem Korporale und einem Kelch in der Kapelle der Burg vorgefunden wurde. Der Gegenstand wird zunächst im Bezugssystem der mit diesem Eintrag verzeichneten Objekte verortet, das Ensemble wiederum dem räumlichen Kontext der Burgkapelle in der Burg Wurmberg zugeordnet. Durch die Angabe der Holzart, nämlich Zypresse, kann das Objekt über die Annotation zur Kategorie Objektmaterial / Holz plausibel eingeordnet werden.<sup>21</sup>

Die in ‚ITEM/inventories‘ bereitgestellten Daten erlauben jedoch auch einen tieferen Einblick in die materielle Beschaffenheit von Objekten, sofern sie in den Inventaren ausgewiesen wurden. Für die insgesamt 48 dem Oberbegriff Holz zugeordneten Objekte sind acht unterschiedliche Holzarten verzeichnet, neben Ahorn, Eiche und Fichte auch exklusivere wie Ebenholz oder eben das bereits genannte Zypressenholz. Die Beschreibung des Holzes im Inventar umfasste bei neun Objekten aus Holz auch die Angabe der Maserung.<sup>22</sup> So fanden die inventarisierenden Personen im Herrenzimmer der Mindelburg einen als *paternoster* bezeichneten Rosenkranz aus Holz oder in jenem der Burg Würnitz eine Truhe, deren Holz eine spezifische, durch tangential geschnittenes Holz entstandene und als Fladerung bezeichnete Maserung aufwies. Auffällig ist bei der Bezeichnung der Maserung, dass alle Gegenstände aus Holz tendenziell als in höhergewerteten

---

19 Das Abfrageergebnis enthält derzeit insofern noch Unschärfen, als in den Quellen das Material von Gegenständen nicht explizit ausgewiesen wurde, wenn seine Verwendung als selbstverständlich erachtet wurde, wie etwa Holz bei Bettgestellen, Tischen oder dergleichen. Hier bietet es sich an, zukünftig manuell und semiautomatisiert nachzubessern.

20 Eine Facetten-Suchabfrage ermöglicht es, Daten anhand mehrerer Filteroptionen zu durchsuchen, vgl. Bleier u. Stoff 2024.

21 Für eine weitere Abfrage, z. B. zusätzlich zu Holzarten nach speziellen Kleidungsstücken im gleichen Raum, muss eine Cypher-Abfrage für Neo4j formuliert werden. Hier endet die einfache Facettensuche mit ‚ITEM‘.

22 Mit der Bedeutung von geflädertem Holz beschäftigen sich derzeit Isabella Nicka und Miriam Landkammer im Rahmen des Projektes ‚Wie das Material ins Bild kam. Kulturelle Innovationen interdisziplinär mit Künstlicher Intelligenz und DH erforschen (KIKI)‘: <https://www.imareal.sbg.ac.at/imareal-projekte/kiki/> (Zugriff: 20.03.2025). Vgl. auch den Beitrag von Isabella Nicka, Miriam Landkammer, Andreas Uhl, Michael Linortner und Johannes Schuiki in diesem Themenheft.

**Thesaurus**

- ▶ Objektbezeichnung
- ▶ Ort
- ▶ Raum
- ▶ Raumbezeichnung
- ▶ Raumtypen

**Facetten**

- ▶ Archivzitate - 9
- ▶ Literaturzitate - 2
- ▶ Objektbezeichnung ...
- ▶ Objektname - 1650
- ▶ Objektbewertung - 33
- ▶ Objekteigenschaft - ...
- ▶ Objektfarbe - 20
- ▶ Herstellungstechnik ...
- ▼ Objektmaterial - 86
  - Bein/Knochen - 2
  - Bernstein - 2
  - Elfenbein - 1
  - Feder - 2
  - Gips - 1
  - Glas - 2
  - Granat - 1
  - Holz - 9
  - Holz/Ahorn - 2
  - Holz/Ebenholz - 1
  - Holz/Eiche - 3
  - Holz/Fichte - 2
  - ☑ Holz/Fladern - 9

**Beschreibungsdaten**

flad

E666  Kopf Objekt

E666 Holz/Fladern Material (Objekt)

E1180  Paternoster Objekt

E1181 Holz/Fladern Material (Objekt)

**Paternoster (Objekt) E1180**

RDF-URI:  
[http://item.imareal.sbg.ac.at/rdf/v1/description/entity/00019015\\_16\\_E1180](http://item.imareal.sbg.ac.at/rdf/v1/description/entity/00019015_16_E1180)

Name Paternoster  
 Eigenname fläderene paternoster  
 Bezeichnung Rosenkranz  
 Hersteller  
 Anzahl 6  
 Begriffstyp historisch  
 Objektgattung  
 Sub ID  
 Post  
 Ante  
 Datierung  
 exakt  
 Objekttyp wobjekt

**Beziehungen**

→ **Eingehende Beziehungen**

**Abschnitt**  
 Mindelburg 003 (E713) ←   
 Herrenzimmer (E74) ←

→ **Ausgehende Beziehungen**

**Material (Objekt)**  
 Holz/Fladern (E1181) →

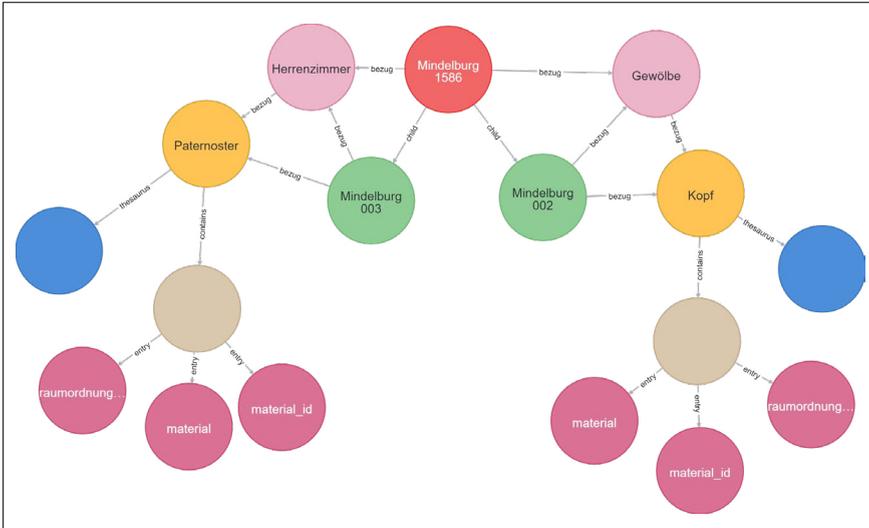
**Abb. 1** | Facettensuche (vergrößert): Mindelburg, Holz/Fladern, Eintrag Paternoster. IMAREAL, Backend-Abfrage ‚ITEM/inventories‘.

Räumlichkeiten befindlich beschrieben wurden.<sup>23</sup> Bewertungen werden auch in anderen Konstellationen sichtbar, etwa wenn in der Kammer der Burg Moos *allerlay znichte plunder* vorgefunden und im Inventar vermerkt wurde oder berichtet wurde, dass die Siedeltruhe in der Herrenstubenkammer der Burg Aistersheim lediglich *plunder* enthielt (Abb. 1).

Diese Anforderungen soll auch das Recherchetool ‚ITEM/accounts‘ erfüllen, das sich derzeit im Aufbau befindet. In den letzten Jahren wurde eine Reihe von digitalen Editionen spätmittelalterlicher und frühneuzeitlicher Rechnungsbücher vorgestellt, die in unterschiedlicher Intensität auf die editorischen Anforderungen der Quellengattungen reagieren.<sup>24</sup> Zwar wurden meist auch Möglichkeiten der inhaltlichen Erschließung diskutiert und über unterschiedliche Abfragesysteme gelöst, in beiden Fällen lehnten diese sich jedoch im besten Fall an die Struktur der Aufzeichnungen an, die sich ihrerseits aus den administrativen Zwecken ergab. So bietet beispielsweise die aktuell veröffentlichte Beta-Version der ‚Jahrrechnungen der Stadt Basel 1535 bis 1610 – digital‘ einen sehr ausgereiften Vorschlag

<sup>23</sup> Zur sozialen Kategorisierung von Räumen auf Burgen vgl. Handzel u. Kührtreiber 2015.

<sup>24</sup> Vgl. Vogeler 2015.



**Abb. 2 |** Ansicht der Graphdaten von allen Objekten mit dem Material gefladertes Holz, die im Inventar der Mindelburg von 1586 aufgelistet wurden. Die Begriffe ‚Paternoster‘ und ‚Kopf‘ werden über die Begriffsfelder ‚Materialität‘ und ‚Burg‘ verbunden. IMAREAL, Backend-Abfrage ‚ITEM/inventories‘.

zur digitalen Edition und Erschließung von Rechnungsbüchern.<sup>25</sup> Sie bietet eine strukturierte Darstellung der Einnahmen und Ausgaben, vor allem aber trägt sie dem bisher kaum gelösten Problem der Berechenbarkeit der Einnahmen und Ausgaben Rechnung. Sucht man jedoch nach Objekten oder Objektgruppen und nach Informationen zu diesen Objekten, die die Schreiber der Rechnungsbücher für aufzeichnungswürdig erachteten, bringt auch die erweiterte Suche in den nach Ämtern und Ausgaben geordneten Rubriken keine befriedigenden Ergebnisse.<sup>26</sup> Eine stichprobenartige Suche nach dem Begriff ‚Holz‘ ergibt für einen Zeitraum von knapp 100 Jahren lediglich fünf Treffer, für eine Suche nach Gegenständen aus Holz müssen Trunkierungen oder reguläre Ausdrücke in Kombination mit Quellenbegriffen bemüht werden. Einen ersten Lösungsansatz bietet hierzu die Edition der Augsburger Baumeisterbücher, die neben einem Namen-, Orts- und Körperschaftsregister auch eine semantische Indizierung in Form eines Schlagwortregisters aufweist.<sup>27</sup> Diese ist alphabetisch nach Überkategorien und Untergruppen erschließbar. Sucht man darin beispielsweise nach dem Begriff ‚Holz‘, wird man in

<sup>25</sup> <https://gams.uni-graz.at/context:srbas?mode=projekt> (Zugriff: 20.03.2025).

<sup>26</sup> Vgl. Burghartz 2015.

<sup>27</sup> Die Augsburger Baumeisterbücher. Digitale Edition der mittelalterlichen Stadtrechnungen 1320 bis 1466. Online: <https://augsburger-baumeisterbuecher.de/> (Zugriff: 20.03.2025).

der Überkategorie ‚Holzarten‘ fündig, während die kurz zuvor gelistete ‚Banktruhe‘ aufgrund einer mangelnden Spezifizierung im Quelleneintrag keinen Bezug zum Material aufweist. Daher listet eine allgemeine Suche nach ‚Holz‘ nur jene Seiten des Rechnungsbuches auf, in denen Holz – meist als Baumaterial – Erwähnung findet, jedoch nicht als Spezifizierung der materiellen Beschaffenheit eines Objektes. ‚ITEM/inventories‘ und ‚ITEM/accounts‘ bieten somit eine Suchmöglichkeit, die keine vertiefte Kenntnis der Quellenbegriffe voraussetzt (Abb. 2).

#### 4 Workflow mit Thesauri und Ontologien: Mapping als Methode zur Datenerschließung

Bisher wurde mit dem Rechnungsbuch des Kremser Stadtbaumeisters Wilpold Harber (entstanden 1457–1459) ein erster *Proof of Concept* für ‚ITEM/accounts‘ durchgeführt. Basierend auf der Edition von Gerhard JARITZ, wurden Digitalisate des Rechnungsbuches in Transkribus entlang der von DEPCHA entwickelten ‚Bookkeeping-Ontology‘ annotiert, um eine übersichtliche Annotationstiefe zu gewährleisten.<sup>28</sup> Von dieser Ontologie wurden 12 Tags (`bk_entry`, `bk_money`, `bk_subtotal`, `bk_from`, `bk_to`, `bk_service`, `bk_commodity`, `bk_when`, `bk_where`, `bk_status`, `bk_account`, `bk_agent`) ausgewählt, um die entsprechenden Einträge im Zuge der Bearbeitung mit Transkribus zu markieren.<sup>29</sup> Für eine tiefere semantische Erschließung von Objekten, Personen, Tätigkeiten, Zeit und Räumen wurden die Tags um textspezifische Attribute erweitert.<sup>30</sup>

Entsprechend der bisher üblichen Vorgehensweise von Transkriptionen wurden im ersten Testlauf von ‚ITEM/accounts‘ zunächst die erwähnten Rechnungsbücher annotiert. Anschließend wurden die Daten in die Neo4j-Graphdatenbank übertragen. Die kontrollierten Vokabulare der Thesauri zur systematischen Erschließung ermöglichen eine standardisierte Beschreibung jedes einzelnen Objekts. Sie geben einen Standard vor, der es ermöglicht, Inventare und Rechnungsbücher für realienkundliche Fragestellungen inhaltlich zu erschließen, ohne zunächst eine Transkription oder Edition anfertigen zu müssen. Mit der Annotation und der Verknüpfung des Vokabulars mit den jeweiligen Begriffen, Wortgruppen oder Abschnitten im Digitalisat werden die Daten generiert, die für Thesaurus-basierte Abfragen benötigt werden. Der Text an sich bleibt bei diesen

<sup>28</sup> Stadtarchiv Krems, Mappe Amts-Rechnungsbücher Krems und Stein 1455–1480; Edition mit Kommentar: Jaritz 1975/1976.

<sup>29</sup> Vgl. Pollin 2020. Die Entscheidung für die Verwendung der ‚Bookkeeping Ontology‘ fiel vor allem aufgrund der nachhaltigen Datensicherung im Bereich der Rechnungsbuchforschung.

<sup>30</sup> Dieser Arbeitsschritt wurde von Roman Plietl durchgeführt, der sich mit den Begrifflichkeiten intensiv auseinandergesetzt und einen gut strukturierten Begriffspoolentwurf vorgelegt hat.

stark strukturierten Quellengattungen für diese Fragestellungen zwar als Kontext relevant und kann über die bereitgestellten Digitalisate oder – im Fall der auf ‚RaumOrdnungen‘ basierenden Datensätze – durch Teiltranskripte erschlossen werden. Die in einem ersten Bearbeitungsschritt erstellten Annotationen und Standardisierungen können für eine spätere Transkription bzw. Edition herangezogen werden. Die Thesauri strukturieren den Text wie auch die Daten bereits stark vor. In einem weiteren Schritt werden die Daten und Thesaurusbegriffe mittels LOD mit Informationen aufgeladen und verknüpft. Der so entstandene *Knowledge Graph*<sup>31</sup> ist graphisch nicht nur visuell übersichtlich (sofern die eingeblendete Knotenanzahl kontrolliert gehalten wird), sondern auch maschinenlesbar. Der Graph liefert Kontext, den *Large Language Models* (LLMs) wie GPT4o<sup>32</sup> von OpenAI mit Hilfe von *Retrieval Augmented Generation* (RAG)<sup>33</sup> dazu nutzen können, um wissensintensive Aufgaben zu lösen.<sup>34</sup> Eine solche kann eine natürlichsprachliche Frage an die Graphdatenbank sein, die normalerweise SPARQL- oder Cypher-Kenntnisse voraussetzt.<sup>35</sup> LOD kann durch die Verknüpfung von Metadaten zu verschiedenen historischen Quellengattungen den Einsatz eines RAG verbessern – sie liefern den Kontext, der nötig ist, um Zusammenhänge zu finden.

Statt einer manuellen Eingabe der Annotationen in ‚ITEM‘ liefert Transkribus mit der Exportfunktion in PAGE XML<sup>36</sup> alle Daten inklusive der Annotationen und Eigenschaften aus. Ein am IMAREAL entwickelter Prototyp eines Import-Tools übersetzte diese schließlich in das bekannte RDF-Schema, das auf der Datenbankarchitektur von ‚REALonline‘ basiert. Damit Daten aus ‚ITEM/accounts‘ und ‚ITEM/inventories‘ gemeinsam abgefragt werden können, wurden die Klassen und Unterklassen der ‚Bookkeeping-Ontology‘ auf die Ontologie der Inventare ‚gemappt‘ – sie wurden schlichtweg übersetzt. Hier zeigt sich

---

31 *Knowledge Graphs* sind Wissensdatenbanken, die als Graphen strukturiert sind. Entitäten werden mit Relationen untereinander verknüpft. Somit liegen alle Daten in Subjekt-Prädikat-Objekt-Form natürlichsprachlich vor. Die vorher genannte *zipresseins tasl* wird als „Tassel hat das Material Holz/Zypresse“ gespeichert.

32 OpenAI 2024. GPT4o ist die aktuelle Version des LLMs GPT (Stand Dezember 2024). Die für *Reasoning* optimierte Version GPTo1 wurde in einer Preview-Fassung ebenfalls für Testzwecke herangezogen.

33 RAG kombiniert ein LLM mit einem Informationsabrufsystem (*Retriever*), das relevante externe Datenquellen wie *Knowledge Graphs*, Datenbanken oder Web-Dokumente durchsucht. Dadurch kann das LLM nicht nur auf sein vorab trainiertes Wissen zugreifen, sondern auch aktuelle und kontextrelevante Informationen in seine Antworten einfließen lassen.

34 Vgl. Feng, Feng, Zhao u. a. 2024.

35 Die *SPARQL Protocol and RDF Query Language* ist eine Abfragesprache für Daten, die in RDF vorliegen. Cypher ist wie Neo4j eine Abfragesprache für Graphdatenbanken, die die Abfrage von Knoten, Beziehungen und deren Eigenschaften in Netzwerken ermöglicht.

36 PAGE XML ist ein Standard zur Codierung von gescannten Dokumenten, der u. a. von Transkribus verwendet wird.

bereits die Ähnlichkeit der beiden Quellengattungen, denn bis auf Ausnahmen wie `bk_service`, eine Dienstleistung, die in einem Inventar nicht vorkommt, konnten die Tags in die bereits vorhandene Struktur übersetzt werden. ‚ITEM‘ als Schnittstelle zwischen diesen Daten ist damit ein erster, vielversprechender Schritt.

Bei der Erschließung unterschiedlicher Überlieferungsformen gehen wir davon aus, dass Text, Bild und materielle Objekte gleichwertige Repräsentationen kultureller Güter darstellen. Die digitale Repräsentation von physisch überlieferten, in Bildern dargestellten oder als Text beschriebenen Objekten erfolgt über sprachliche Klassifikationen in Form hierarchisch strukturierter Thesauri und Ontologien, da es mit ihnen möglich ist, über eine gemeinsame Abfrageschnittstelle Schnittmengen an Ergebnissen zu erzielen. Für die Erschließung von Inventaren und Rechnungsbüchern wurde dabei auf die bereits bestehende Ontologie der Bilddatenbank ‚REALonline‘ aufgebaut und diese um die Spezifika der Quellengattung erweitert. Damit kann für jeden Datensatz (zum Beispiel Inventar oder Rechnungsbuch) eine beliebige Zahl an Entitäten erfasst werden, die einem Set an Entitätstypen zugeordnet sind.<sup>37</sup> Darüber hinaus können Informationen zu einzelnen Bestandteilen und Bezüge („ist Teil von“) erfasst werden.

Somit dient nicht nur die ‚REALonline‘-Ontologie<sup>38</sup> als Basis, sondern auch die zugehörigen Thesauri, die über 1,2 Millionen Begriffe umfassen. Die zusätzlichen Verknüpfungsmöglichkeiten mit der Bilddatenbank des IMAREAL erscheinen für zukünftige Projekte vielversprechend. Das für ‚ITEM‘ benötigte Vokabular der Thesauri basiert auf dieser für die Erschließung der Bilddaten genutzten Grammatik bzw. auf der Ontologie (Subjekt, Prädikat, Objekt nach RDF) und wird nach Bedarf erweitert.

Ontologie und Thesauri werden aufgrund der Heterogenität der Quellen ‚gemappt‘ – mit der Möglichkeit, dieses Mapping auch um Thesauri zu erweitern, die für andere Quellengattungen entwickelt wurden. Durch das Mapping werden nicht nur semantische Lücken geschlossen, sondern es wird auch die Komplexität historischer Quellengattungen reduziert, was wiederum gezielte Abfragen und vergleichende Analysen erleichtert. LOD bedeutet hier somit nicht nur die Verknüpfung mit kontrollierten Vokabularen – auch wenn dies allein schon äußerst nützlich ist –, sondern auch die Verknüpfung von unterschiedlichen Daten und Projekten der Mittelalterforschung am IMAREAL, um eine systematische Abfrage verschiedener Quellengattungen zu ermöglichen. Ein Mapping sieht dabei auf semantischer Ebene so aus, dass einzelne Entitäten dann als zusammengehörig markiert werden, wenn sie ausreichend semantische Ähnlichkeit aufweisen. Objektbezeichnungen der Inventare können mit Gegenständen (`bk_commodity`)

---

<sup>37</sup> Vgl. Matschinegg u. Nicka 2018, S. 14.

<sup>38</sup> Vgl. Matschinegg u. Nicka 2024, S. 210–214.

genauso problemlos verbunden werden wie Künstler\*innen (aus ‚REALonline‘), Personen aus ‚ITEM/inventories‘ und Namen/Gruppen (bk\_party) aus ‚ITEM/accounts‘. Darüber hinaus ist auch ein Mapping auf Thesaurus-Ebene möglich: Die Bezeichnung ‚Altarkreuz‘ kommt in allen drei Datensätzen vor. Dieses Mapping befindet sich in Entwicklung und Erprobung.

Nach diesem Prozess wird die Annotation schließlich in die Neo4j-Graphdatenbank mit je nach Datenquelle austauschbarem Importskript importiert. Das Web-Frontend dient als *User Interface* für etwaige Forschungsfragen mittels der Facettensuche, wodurch User\*innen für einfache Abfragen nicht mit Cypher-Abfragen im Neo4j-Browser arbeiten müssen. Sie können für eine komplexere Suche weiterhin auf die Expert\*innensuche zurückgreifen.<sup>39</sup> Durch das *Workflow Engineering* wurden Vorbereitungen getroffen, diesen Schritt nicht manuell über das Frontend bearbeiten zu müssen, sondern eine Suche automatisiert von spezialisierter Software zu ermöglichen. In dem von uns skizzierten Fall ist das der Einsatz einer spezialisierten RAG, durch die LLMs mit Expert\*innenwissen ausgestattet werden. In der Folge können komplexe Cypher-Kommandos erstellt, automatisiert an den nächsten Agenten weitergegeben und ausgewertet werden. Dieser Workflow ermöglicht es, natürlichsprachliche Fragen zu formulieren und das strukturierte Datenmaterial damit auszuwerten. Die Komplexität macht eine entsprechende Infrastruktur erforderlich.

Bisher besteht die Arbeit an ‚ITEM/inventories‘ und ‚ITEM/accounts‘ vor allem im *Workflow Engineering*. Das bedeutet: Jeder Arbeitsschritt des Arbeitsprozesses wurde so konzipiert, dass er von verschiedenen Personen oder *Agents* durchgeführt werden kann. Diese *Agents* können auch semi-automatisierte Tools sein, wie beispielsweise eine für diesen Arbeitsprozess spezialisierte generative KI.<sup>40</sup> Der Workflow für eine KI-basierte Unterstützung ist vorbereitet und wurde bereits getestet.<sup>41</sup>

In ersten Versuchen wurde mit einer bestehenden Transkription gearbeitet. Zeitsparender und für realienkundliche Fragestellungen ausreichend erscheint hingegen eine Annotation des Digitalisats ohne vorangehende Transkription. Annotationen können im *Stand-off*-Verfahren separat angefertigt werden.<sup>42</sup> Durch die Anreicherung mit Meta-Informationen im RDF lassen sich zusätzlich die Koordinaten der Textstelle im Digitalisat angeben. Somit wird ein eigenes Annotationstool implementiert.<sup>43</sup> Diese Meta-Informationen sind zum Beispiel

---

<sup>39</sup> Vgl. Matschinegg u. Nicka 2024, S. 216–219.

<sup>40</sup> Vgl. Wu, Bansal, Zhang u. a. 2024.

<sup>41</sup> Erste Ergebnisse werden Anfang 2025 als Webapplikation veröffentlicht.

<sup>42</sup> Als *Stand-off*-Verfahren wird die Annotation von Primärdaten bezeichnet, wenn diese getrennt von den Primärdaten gespeichert werden. Vgl. dazu Klug 2021.

<sup>43</sup> Vgl. Vogeler 2022.

in PAGE XML mit den Tags <TextRegion>, <TextLine> und <Coords> dem Text zugeordnet. Durch die zukünftige Unterstützung von Tools, wie sie für die Bildannotation bei ‚REALonline‘ schon existieren, wird der Arbeitsprozess erleichtert, da die Passagen nur noch mit entsprechenden Rahmen markiert werden müssen. Durch die Übersetzung in ein einheitliches Datenformat bleibt die weiterführende Verwendung für andere Fragestellungen möglich. Im Speziellen erleichtert dies die Schnittstellenfunktion von ‚ITEM‘, da mehr Personen mit den Daten arbeiten können, gleichzeitig die Annotation vereinfacht wird und auf Open-Source-Software im Prozess gesetzt werden kann.

## 5 Fazit

Das am IMAREAL entwickelte digitale Abfrage-Tool ‚ITEM‘ dient als Schnittstelle, um aus den Datenbanken ‚ITEM/accounts‘ und ‚ITEM/inventories‘ gezielt Rechnungsbuch- und Inventardaten hinsichtlich Begriffen aus den *Material Culture Studies* –wie im vorgestellten *Use Case ‚Holz‘*– abfragen zu können. Dies ist durch die Formulierung von Suchabfragen in neuhochdeutscher Sprache möglich. Die historischen Begrifflichkeiten werden über die am IMAREAL entwickelten strukturierten Thesauri erschlossen. ‚ITEM‘ ermöglicht es daher, heterogene Quellengattungen wie Inventare und Rechnungsbücher gemeinsam zu analysieren. Der Zugriff der digitalen Tools des IMAREAL auf die durch Annotation von Bildquellen, Inventaren und Rechnungsbüchern erstellten und aufeinander abgestimmten Thesauri bietet eine standardisierte, zugängliche Abfragemöglichkeit einer Vielzahl von Quellen und Quellengattungen. Dieser interdisziplinäre Ansatz stärkt nicht nur die Forschung im Bereich der *Material Culture Studies*, sondern schafft auch eine Grundlage für zukünftige semantische Analysen und *Linked Open Data*-Projekte. Damit leistet ‚ITEM‘ einen entscheidenden Beitrag zur Etablierung neuer Standards in der digitalen Mediävistik und ermöglicht innovative Forschungsansätze, die über die Grenzen einzelner Disziplinen hinausgehen.

## Literaturverzeichnis

- Antenhofer, Christina:** Die Familienkiste: Mensch-Objekt-Beziehungen im Mittelalter und in der Renaissance (Mittelalter-Forschungen 67). Ostfildern 2021.
- Antenhofer, Christina, Elisabeth Gruber-Tokić, Gerald Hiebel u. a.:** Inventarisierung als Praxis. Die semantischen Welten der Burgeninventare des historischen Tirols. In: Christina Antenhofer u. Richard Schober (Hgg.): *Tiroler Heimat 87* (2023), S. 81–100.
- Antenhofer, Christina u. Ingrid Matschneegg:** Inventare als Korrektiv des „männlichen Blicks“ auf Burgen. Eine

- Spurensuche entlang italienischer Brautschatzinventare und der Inventare von Schloss Bruck bei Lienz. In: Österreichische Zeitschrift für Geschichtswissenschaften 32 (2022), S. 165–178. <https://doi.org/10.25365/oezg-2021-32-3-9> (Zugriff: 20.03.2025).
- Christina Antenhofer, Ingrid Matschinegg, Walter Brandstätter u. a.:** Hohen-salzburg digital. Historische Daten zur materiellen Raumausstattung und Raumnutzung erschließen und verlinken. In: Jahrbuch für Hausforschung 5 (2023), S. 144–156.
- Bleier, Roman u. Sebastian Stoff:** Inhalte explorativ durchsuchen. Geisteswissenschaftliche Forschungsdaten und Facet-tendrilldown. In: Victoria Kumar, Gerald Lamprecht, Lukas Nievöll u. a. (Hgg.): Erinnerungskultur und Holocaust Education im digitalen Wandel. Bielefeld 2024, S. 211–228.
- Burghartz, Susanne (Hg.):** Jahrrechnungen der Stadt Basel 1535 bis 1610 – digital. <https://gams.uni-graz.at/context:srbas> (Zugriff: 20.03.2025).
- Eide, Oyvind, Christopher Pollin u. Georg Vogeler:** bookkeeping. 2015. <https://github.com/GVogeler/bookkeeping> (Zugriff: 20.03.2025).
- Feng, Zhangyin, Xiaocheng Feng, Dezhi Zhao u. a.:** Retrieval-Generation Synergy Augmented Large Language Models. In: ICASSP 2024 – 2024 IEEE International Conference on Acoustics, Speech and Signal Processing. 2024, S. 11661–11665. <https://doi.org/10.1109/ICASSP48485.2024.10448015> (Zugriff: 20.03.2025).
- Gleba, Gudrun u. Niels Peterson:** *int erste*. Zur Einleitung. In: Dies. (Hgg.): Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit. Formen und Methoden der Rechnungslegung: Städte, Klöster, Kaufleute. Göttingen 2015, S. 7–11.
- Gruber, Elisabeth:** *Raittung und außgab zum gepew*. Kommunale Rechnungspraxis im oberösterreichischen Freistadt. Edition und Kommentar der Stadtgrabenrechnung (1389–1392) (Quelleneditionen des Instituts für Österreichische Geschichtsforschung 13). Wien 2015.
- Gruber, Elisabeth:** *Umb ain polsterziechen in das rathaus darauf der burgermaister sitzet*. Dinge und Objekte in der städtischen Rechnungsüberlieferung, in: Sabine von Heusinger u. Susanne Wittekind (Hgg.): Die materielle Kultur der Stadt im Spätmittelalter (45. Frühjahrskolloquium des Instituts für vergleichende Städtegeschichte und des Kuratoriums für vergleichende Städtegeschichte e. V. 2019). Wien, Köln 2019, S. 39–50.
- Gruber, Elisabeth u. Thomas Kühnreiber:** Menschen – Dinge – Beziehungen. Weinbau als kulturelle Praxis. In: Österreich in Geschichte und Literatur (mit Geographie) 65 (2021), S. 416–427.
- Gruber-Tokić, Elisabeth, Gerald Hiebel, Brigit Danthine u. a.:** Von historischen Texten zu Karten und semantischen Netzwerken. In: Thomas Weinhold (Hg.): 22. Internationale Geodätische Woche Obergurgl 2023. Berlin 2023, S. 11–22.
- Handzel, Josef, Gabriele Klug u. Christina Schmid:** ‚Schwarze Löcher‘: Fragmentierte Räume – Räumliche Fragmente. In: Christian Gastgeber, Christine Gläßner, Kornelia Holzner-Tobisch u. a. (Hgg.): Fragmente. Der Umgang mit lückenhaften Quellenüberlieferungen in der Mittelalterforschung. Wien 2010, S. 185–204.
- Handzel, Josef u. Thomas Kühnreiber:** Herrenstube und Frauenzimmer. Sozial konnotierte Lebensräume auf Burgen im

Kontext der schriftlichen und bauhistorischen Überlieferung am Beispiel von Burg Pürnstein, Oberösterreich. In: Christina Schmid, Gabriele Schichta, Thomas Kühnreiter u. a. (Hgg.): Raumstrukturen und Raumausstattung auf Burgen in Mittelalter und Früher Neuzeit. Heidelberg 2015, S. 530–541.

**Handzel, Josef, Gabriele Schichta u.**

**Christina Schmid:** RaumOrdnungen. Raumfunktionen und Ausstattungsmuster auf Adelssitzen im 14. bis 16. Jahrhundert. In: Christina Schmid, Gabriele Schichta, Thomas Kühnreiter u. a. (Hgg.): Raumstrukturen und Raumausstattung auf Burgen in Mittelalter und Früher Neuzeit. Heidelberg 2015, S. 15–66.

**Jaritz, Gerhard:** Die Rechnungen des Kremser „Stadtbaumeister“ Wilpold Harber aus den Jahren 1457 bis 1459. In: Mitteilungen des Kremser Stadtarchivs 15/16 (1975/1976), S. 1–62.

**Klug, Helmut W.:** Stand-off-Markup. In: Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner (Hgg.), KONDE Weißbuch. HRSM Projekt „Kompetenznetzwerk Digitale Edition“ 2021. Handle: hdl.handle.net/11471/562.50.171 (Zugriff: 20.03.2025)

**Kühnreiter, Thomas:** ‚Raum-Ordnungen‘. Raumfunktionen und Ausstattungsmuster auf Adelssitzen im 14.–16. Jahrhundert. Ein Forschungsprojekt am ‚Institut für Realienkunde‘ (Medium Aevum Quotidianum 55). Krems 2005, S. 59–79.

**Kühnreiter, Thomas:** Dinge ordnen. Adeliges Haushalten in der Frühen Neuzeit am Beispiel der Herrschaft Windhag, Gemeinde Windhaag, Oberösterreich. In: Institut für Realienkunde des Mittelalters und der frühen Neuzeit (Hg.): Object Links – Dinge in Beziehung. Wien, Köln, Weimar 2019, S. 43–73.

**Kühnreiter, Thomas u. Heike Schlie:**

Holz als Geschichtsstoff. Das Materielle in den Dingkulturen. In: MEMO 1 (2017), S. 1–11. <http://dx.doi.org/10.25536/20170101> (Zugriff: 20.03.2025).

**Matschinegg, Ingrid:** Inventarisierte Objekte verlinken: Schreibzeug, Schreibtisch, Schreibstube. Objekte und Orte des Schreibens auf Burgen (15. und 16. Jahrhundert). In: Institut für Realienkunde des Mittelalters und der frühen Neuzeit (Hg.): Object Links – Dinge in Beziehung. Wien, Köln, Weimar 2019, S. 75–94.

**Matschinegg, Ingrid und Isabella Nicka:** REALonline Enhanced. Die neuen Funktionalitäten und Features der Forschungsbilddatenbank des IMAREAL. In: MEMO 2 (2018), S. 10–32. <http://dx.doi.org/10.25536/20180202> (Zugriff: 20.03.2025).

**Matschinegg, Ingrid u. Isabella Nicka:** Visuelle Medien mit digitalen Methoden erforschen. In: Christina Antenhofer, Christoph Kühberger u. Arno Strohmeier (Hgg.): Digital Humanities in den Geschichtswissenschaften. Wien 2024, S. 205–225.

**MEMO 1 (2017):** Holz in der Vormoderne. Werk-Stoff, Wirk-Stoff, Kunst-Stoff. <http://dx.doi.org/10.25536/2523-2932012017>.

**OpenAI:** Hello GPT-4o. <https://openai.com/index/hello-gpt-4o/> (Zugriff: 20.03.2025).

**Pollin, Christopher:** Digital Edition Publishing Cooperative for Historical Accounts and the Bookkeeping Ontology. In: Thomas Riechert, Francesco Beretta u. George Theodore Bruseker (Hgg.): Proceedings of the Doctoral Symposium on Research on Online Databases in History. 2019, S. 7–14.

<https://ceur-ws.org/Vol-2532/paper1.pdf>  
(Zugriff: 20.03.2025).

**Pollin, Christopher:** Digitale, formale Methoden und Modelle in den Geschichtswissenschaften. Am Beispiel digital editierter historischer Rechnungsbücher. Master Thesis. Graz 2020.

**Pollin, Christopher:** RDF, RDFS, OWL. In: Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner (Hgg.), KONDE Weißbuch. HRSM Projekt „Kompetenznetzwerk Digitale Edition“ 2021. Handle: hdl.handle.net/11471/562.50.131. PID: o:konde.131 (Zugriff: 20.03.2025).

**Radkau, Joachim:** Holz. Wie ein Naturstoff Geschichte schreibt. München 2007.

**Rippmann, Dorothee:** Leben, Arbeit und materielle Kultur im Lichte pragmatischer Schriftlichkeit in der Schweiz. In: Gudrun Gleba u. Niels Peterson (Hgg.): Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit. Formen und Methoden der Rechnungslegung: Städte, Klöster, Kaufleute. Göttingen 2015, S. 210–254.

**Simon-Muscheid, Gabriele:** Die Dinge im Schnittpunkt sozialer Beziehungsnetze. Reden und Objekte im Alltag. Göttingen 2004.

**Vogeler, Georg:** Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert? In: Constanze Baum u. Thomas Stäcker (Hgg.): Grenzen und Möglichkeiten der Digital Humanities (Sonderband der Zeitschrift für digitale

Geisteswissenschaften 1). 2015. [http://dx.doi.org/10.17175/sb001\\_007](http://dx.doi.org/10.17175/sb001_007) (Zugriff: 20.03.2025).

**Vogeler, Georg:** Reading an XML Text Like a Human with Semantic Web Technologies. Learning from the Basel City Accounts as Digital Edition. In: Tina Asmussen, Eva Brugger, Maike Christadler u. a. (Hgg.): Materialized Histories. Eine Festschrift 2.0 für Prof. Dr. Susanna Burghartz. Basel 2021. <https://mhstories.hypotheses.org/783> (Zugriff: 20.03.2025).

**Vogeler, Georg:** Edition – Protoedition – Reproduktion. Der digitale Wandel. In: Geschichte in Wissenschaft und Unterricht 73 (2022), S. 498–511. [https://elibrary.utb.de/doi/10.5555/gwu-9%2B10-2022\\_02](https://elibrary.utb.de/doi/10.5555/gwu-9%2B10-2022_02). (Zugriff: 20.03.2025).

**Wu, Quingyun, Gagan Bansal, Jieyu Zhang u. a.:** AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. 2024. <https://doi.org/10.48550/arXiv.2308.08155> (Zugriff: 20.03.2025).

**Wunschel, Andreas:** Aus dem Wald zum Holzhandwerker in das tägliche Leben. Wie hölzerne Gegenstände das Mittelalter prägen. In: LWL-Freilichtmuseum Hagen (Hg.): Echt alt! Mittelalterliches Handwerk ausgegraben (Forschungsbeiträge zu Handwerk und Technik 33). Hagen 2018, S. 78–93.

# Bischöfe im Graph

## Zur Transformation großer Datenbestände in eine graphbasierte Umgebung

### Kontakt

**Jun.-Prof. Dr. Matthias Weber,**  
Ruhr-Universität Bochum,  
Historisches Institut,  
Universitätsstraße 150,  
44801 Bochum,  
matthias.weber@rub.de  
 <https://orcid.org/0000-0003-2198-8989>

**Abstract** The transfer of research data into digital environments offers the possibility of opening up these data in new contexts, addressing new research questions, and enabling new queries within these data. This process is currently being implemented in an exemplary manner with data that was collected during preliminary work on the ‘Regesta Imperii’ academy project. The focus is on German bishops between 1024 and 1125 in their relationships with kingship and papacy. The aim of the project is to transform the data currently available in a text file into a graph-based environment. This article provides information about the source material, the necessary preparatory work in the form of a conversion of the material into a tabular form for importing into the graph database, and the initial results of the transformation.

**Keywords** Bishops; Digital Medieval Studies; Graph-Based Database; Interlinked Research Data

## 1 Einleitung

Die Rolle der Bischöfe im europäischen Mittelalter kann nicht hoch genug eingeschätzt werden. Sie fungierten als Verwalter ihrer Kirchengemeinde und Seelsorger, als Stadtherren und Territorialfürsten, als Heerführer sowie in Reichsangelegenheiten für Könige und Kaiser. Ihre Bischofsstädte waren, insbesondere im deutschen Reich, regelmäßige Aufenthaltsorte der Herrscher, die sich dort zu hohen Festtagen oder Reichstagen aufhielten, Urkunden

ausstellten und Recht sprachen. Die Position des Episkopats kann treffend mit den Worten von John S. OTT und Anna Trumbore JONES bilanziert werden: „The bishop was unquestionably one of the most important individuals of the European Middle Ages. [...] He did not simply stand at the center of things – he was the center.“<sup>1</sup> Entsprechend der ihnen zugesprochenen Bedeutung sind den Bischöfen kaum mehr zu überblickende Forschungen gewidmet worden, die sich neben der in der Spätantike vollzogenen Genese des Amtes vor allem mit dessen Ausformungen unter den frühen Herrscherdynastien der Merowinger, Karolinger, Ottonen und Salier auseinandergesetzt haben.<sup>2</sup> Insbesondere in der Zeit der zwischen 1024 und 1125 regierenden Könige und Kaiser der salischen Dynastie standen die Bischöfe im Mittelpunkt eines zwischen König- und Papsttum ausgetragenen Konfliktes, dessen vielfältige Facetten unter dem Begriff des Investiturstreits subsumiert werden.<sup>3</sup> Die Bischöfe sahen sich herausgefordert, in dieser Auseinandersetzung Partei zu ergreifen, waren viele von ihnen doch zuvor durch den deutschen König in ihr Amt investiert worden, während andere, insbesondere seit dem Pontifikat Gregors VII. (1073–1085), sich den reformerischen Positionen des Papsttums verschrieben hatten, das die Bischofsinvestitur durch Laien strikt zurückwies. Kurz gesagt, standen die Bischöfe in engster Beziehung zum König und/oder zum Papst, was sich auch anhand einer reichhaltigen Quellenüberlieferung dokumentieren lässt.

Gegenstand des vorliegenden Beitrags ist die Transformation der Forschungsdaten zu den Bischöfen des deutschen Reiches nördlich der Alpen in eine graphbasierte Umgebung sowie die Vorarbeiten, die dafür notwendig sind. Nach einer kurzen Vorstellung des Ausgangsmaterials werden die bisher geleisteten Vorbereitungen und die dabei entstandenen Herausforderungen skizziert, um die Daten für eine digitale Umsetzung aufzubereiten. Anschließend werden erste Ergebnisse vorgestellt.

## 2 Das Ausgangsmaterial

Die anfängliche Intention bei der Zusammenstellung des Quellenmaterials zu den Bischöfen der salischen Zeit war es, für die damals laufende Bearbeitung der Regesten Heinrichs IV.<sup>4</sup> (1056–1106) im Rahmen des Akademieprojektes der

---

1 Ott u. Jones 2007, S. 1 (Hervorhebung im Original).

2 Vgl. exemplarisch aus der reichhaltigen Forschung zu Genese, Ausprägung und Wirkung des Episkopats in der Zeit der genannten Dynastien Scheibelreiter 1983; Patzold 2008; Bode 2015; Zielinski 1984.

3 Zum Investiturstreit vgl. die Überblicksdarstellungen von Zey 2017 und Johrendt 2018.

4 Vgl. Lubich 2016a; Lubich 2016b; Lubich 2018.

‚Regesta Imperii‘ sowie für die anstehende (Neu-)Bearbeitung der Regesten Konrads II. (1024–1039), Heinrichs III. (1039–1056) und Heinrichs V. (1106–1125) alle von diesen Herrschern durchgeführten Investituren im deutschen Reich nachvollziehen zu können. Um dieses Ziel zu erreichen, wurden zunächst alle relevanten Passagen weitgehend zeitgenössischen Quellenmaterials gesammelt, das zu den Investituren der Bischöfe Auskunft gibt. Bald jedoch entwickelte sich das Projekt, auch im Hinblick auf die Ansprüche der ‚Regesta Imperii‘, weiter, da nicht nur die Investitur selbst, sondern jegliche Interaktion zwischen Bischöfen und Königen eine für das Akademievorhaben zu berücksichtigende Handlung darstellt. Schließlich wurde beschlossen, auch die personale Beziehung der Bischöfe zu den Päpsten zu berücksichtigen, um valide Aussagen über deren Handlungen treffen zu können.<sup>5</sup> Zudem gerieten dadurch auch die Beziehungen der Bischöfe untereinander in den Blick, womit zugleich dem aktuellen Forschungsansatz Rechnung getragen wird, die Bischöfe nicht allein in ihrer Rolle als Reichsbischöfe im Dienst für den König zu sehen, sondern sie verstärkt in ihren Funktionen innerhalb ihrer Diözese zu betrachten, wofür eben auch die unabhängig vom König gepflegten Beziehungen der Bischöfe zu anderen Diözesen und deren Bischöfen von Interesse sind.<sup>6</sup> Dabei wurde auch das herangezogene Quellenmaterial erweitert: Gesichtet wurden ein Großteil der im Hoch- und Spätmittelalter entstandenen Texte, wobei historio- und hagiographische Zeugnisse sowie Gedenküberlieferung berücksichtigt wurden.<sup>7</sup> Es entstanden damit Artikel mit von Bischof zu Bischof sehr unterschiedlicher Länge, die zunächst die Erhebungsumstände des Kandidaten sowie seine Sterbedaten nennen, um anschließend seine personalen Beziehungen aufzufächern. Exemplarisch soll dies anhand des Bistums Augsburg vorgeführt werden. Der letzte unter den Saliern erhobene Bischof ist Hermann von Vohburg, der 1096 oder 1097 ins Amt gelangte und bis 1133 amtierte. Sein Artikel beginnt entsprechend mit seiner Investitur, es folgen Angaben zu seiner Weihe und schließlich zu seinem Tod:<sup>8</sup>

---

5 Eine wichtige Vorlage stellten die in mehreren Bänden erschienenen ‚Series episcoporum‘ dar; vgl. Engels u. Weinfurter 1982; Engels u. Weinfurter 1984; Watt 1991; Kluger 1992. Das Projekt ist unvollendet geblieben und umfasst im deutschen Raum einzig die Erzdiözesen von Hamburg-Bremen und Köln. Zur ursprünglichen Intention des Projektes vgl. Weinfurter 1986.

6 Zum geforderten Perspektivwechsel vom gut erforschten Reichsbischof zum wenig beachteten Diözesanbischof vgl. Bihrer u. Bruhn 2019; Bihrer u. Röckelein 2022.

7 Das herangezogene Quellenmaterial basiert auf dem von der Bayerischen Akademie der Wissenschaften digital bereitgestellten Repertorium ‚Geschichtsquellen des deutschen Mittelalters‘, <https://geschichtsquellen.de/start> (Zugriff: 07.11.2024). Eine vollständige Quellen-sichtung ist dadurch zwar noch immer nicht erreicht, aber ein Großteil des bestehenden Quellenmaterials konnte berücksichtigt werden.

8 Im Folgenden: Beginn des Artikels zum Augsburger Bischof Hermann (Quelle: Matthias Weber).

**Hermann** (von Vohburg)

Investiert Dezember 1096 oder Anfang 1097 durch Heinrich IV., geweiht im Jahr 1099 durch Patriarch Ulrich von Aquileja.<sup>166</sup> Gestorben am 19. März 1133.<sup>167</sup>

Die Quellenbelege zu diesen kurzen und prägnanten Angaben sowie Hinweise auf die Forschung sind in den Anmerkungen zu finden. Diese sind im Vergleich zum Haupttext deutlich umfangreicher – nicht zuletzt, da teils ausführliche lateinische Zitate integriert wurden. Die Quellen selbst sind chronologisch angeordnet. Zur Investitur Hermanns von Vohburg ergibt sich damit eine ausführliche Fußnote, die zunächst einen Überblick über die einschlägige Forschung gibt, die sich mit der Erhebung Hermanns zum Bischof von Augsburg im Speziellen oder den Bischöfen der Stadt im Allgemeinen auseinandergesetzt hat. Für die Rekonstruktion des Sterbedatums eines Bischofs wird zusätzlich jede Überlieferung herangezogen, die das Todesjahr und möglichst auch den Todestag zu bestimmen erlaubt:<sup>9</sup>

<sup>166</sup> Vgl. BÖHMER–LUBICH [RI III,2,3 (4)] 99 n. 1399; BONIN, Besetzung, 115; EBELING, Bischöfe I, 8; FEDALTO, Hierarchia, 475; GAMS, Series, 258; GATZ, Bistümer, 67; GEBELE, Hermann, 14; HAUCK, Kirchengeschichte 3, 983; HORN, Augsburg, 260–261; KÄUPER, Verdun, Konstanz und Augsburg, 306–307; KREUZER, Hermann, 2162; MEYER VON KNONAU, Jbb. 4, 479–480; MIEDANER, Verzeichnis, 1202; MOOYER, Verzeichnisse, 3; PAULUS, Bischof Hermann von Augsburg, 91–92; VOLKERT, Hermann, 4; VOLKERT–ZOEPLF, Regesten 1, 221 n. 366, 230 n. 371; WEITLAUFF, Hermann, 30; WENNER, Metropolit, 35; ZOEPLF, Augsburger Bischöfe im Investiturstreit, 316–318; ZOEPLF, Hermann, 630; ZOEPLF, Bistum Augsburg, 109. Zur Bedeutung der Stadt Augsburg für Heinrich IV. vgl. KOTTJE, Bischofsstädte, 148–150. – Über die Nachfolgeregelung der Augsburger Kirche im **Jahr 1096**: Udalschalk, De Eginone et Herimanno c. 12 [Anfang 12. Jh.] (SS 12, 436–437) mit besonderer Berücksichtigung der unrechtmäßigen Einsetzung durch die Hand eines exkommunizierten Laien und dem Vorwurf der Simonie: *sedes [...] constringitur*. – **Weihe**: Udalschalk, De Eginone et Herimanno c. 12 [Anfang 12. Jh.] (SS 12, 437): *Siquidem [...] ordinatur*. Eine später dieser Passage hinzugefügte Notiz, welche die Erhebung Hermanns in das Jahr 1103 verlegt, ergänzt den Fortgang Hermanns um seine durch mehrere Päpste erlittenen Verbote und Exkommunikationen: *Interea [...] excommunicatus*;

<sup>9</sup> Im Folgenden: Fußnote mit Informationen zu Forschungsliteratur und Quellenüberlieferung zur Investitur Hermanns von Augsburg. Die lateinischen Zitate sind an dieser Stelle teils gekürzt wiedergegeben (Quelle: Matthias Weber).

Herbord, Vita Ottonis III, 38 [Mitte 12. Jh.] (SS 20, 766): *Cernitis [...] dicebat*. – Zum **Jahr 1102**: Chronik von der Gründung der Stadt Augsburg bis zum Jahr 1469 [Mitte 15. Jh.] (FRENSDORFF, 302): *ward zû Augspurg bischoff Hermannus*. – **Genannt**: Series episcoporum Augustanorum I, Series ss. Udalrici et Afrae [Mitte 11. Jh.] (SS 13, 335): *Herimannus*; IIa, Series Scheftlariensis [Ende 11. Jh.] (SS 13, 335): *Hermannus episcopus*; III, Series Augustana [Mitte 12. Jh.] (SS 13, 335): *Herimannus*; IV, Series Altahensis [13. Jh.] (SS 13, 335): *Hermannus episcopus*; Series episcoporum Germaniae. Series episcoporum Augustanorum [12. Jh.] (SS 15,2, 1308): *Herimannus*; Chronicon breve episcoporum Augustensium et abbatum s. Afrae [Ende 12. Jh.] (SS 14, 559): *Herimannus*.

Die Erhebung der Amtsdaten der einzelnen Bischöfe ist vergleichsweise leicht umzusetzen. Anspruchsvoller ist es hingegen, die personalen Beziehungen der Bischöfe zu Papst- und Königtum zu rekonstruieren, auf denen der Schwerpunkt der Artikel liegt. Im Anschluss an die Informationen zu Erhebung und Tod folgen in chronologischer Reihung die Beziehungen des Bischofs zu den anderen Personen, in unserem Beispiel zum Patriarchen von Aquileja, der Hermann von Augsburg die Weihe erteilte, zu Papst Paschalis II. (1099–1118), zu dem Hermann erste briefliche und persönliche Kontakte aufgenommen hatte, sowie zu Kaiser Heinrich IV.:<sup>10</sup>

Hermann war zum Zeitpunkt seiner Bischofserhebung exkommunizierter Laie, seine Weihe erhielt er erst 1099.<sup>168</sup> Seine Weihe empfing er von Ulrich von Eppenstein, dem gebannten Patriarchen von Aquileja.<sup>169</sup> April 1099 hielt sich Hermann gemeinsam mit dem Kaiser in Regensburg auf.<sup>170</sup> Ende des Jahres richtete Hermann Briefe an Papst Paschalis II.,<sup>171</sup> auf die dieser im Folgejahr antwortete<sup>172</sup> sowie zwischen 1100 und 1102 einen weiteren Brief an Hermann adressierte.<sup>173</sup> 1102 feierte Hermann in Mainz das Weihnachtsfest gemeinsam mit dem Kaiser<sup>174</sup> und geleitete in seinem Auftrag den zum Bischof von Bamberg investierten Otto Anfang 1103 in dessen Stadt.<sup>175</sup>

Bereits vor der Investitur gepflegte Beziehungen sind dabei nicht berücksichtigt worden. Jede einzelne Handlung wird in der nachgestellten Anmerkung in Bezug gesetzt zu der oder den darüber Auskunft gebenden Quelle(n) und es werden knappe Hinweise zu Forschungsliteratur gegeben:<sup>11</sup>

10 Im Folgenden: Erste Informationen zur Amtszeit Bischof Hermanns von Augsburg hinsichtlich seiner Beziehung zu König- und Papsttum (Quelle: Matthias Weber).

11 Im Folgenden: Fußnoten mit Belegstellen zu den zuvor zitierten ersten Informationen aus der Amtszeit Bischof Hermanns von Augsburg hinsichtlich seiner Beziehung zu König- und

<sup>168</sup> Vgl. dazu HORN, Augsburg, 261; VOLKERT-ZOEPFL, Regesten 1, 230 n. 371. Erzbischof Ruthard von Mainz hatte Hermann (1096 Dezember 4–1098) die Erlaubnis erteilt, sich von irgendeinem katholischen Bischof weihen zu lassen (STIMMING, Mainzer UB 1, 300 n. 396).

<sup>169</sup> Die Weihe übernahm somit nicht der eigentlich vorgesehene, jedoch aufgrund seines Konfliktes mit Heinrich IV. aus seiner Stadt verbannte Metropolit Ruthard von Mainz, sondern der kaisertreue gebannte Patriarch von Aquileja. Dieser musste jedoch zusätzlich durch einen falschen Eid, dargebracht durch einen einfachen Kleriker der Augsburger Kirche, welcher sich als hoher Augsburger Ministerialer ausgab, davon überzeugt werden, dass Ruthard ihm die Erlaubnis erteilt hatte, Hermann zu weihen. Die Weihe fand 1099 nach April 30 statt, Udalschalk, De Eginone et Herimanno c. 12 [Anfang 12. Jh.] (SS 12, 437): *Siquidem [...] fefellit*. Bezüglich der Erlangung der Weihe hatte sich Hermann zuvor in einem undatierten, wohl nach 1098 zu legenden Schreiben an den Mainzer Erzbischof Ruthard gewandt, Monumenta episcopatus Augustani n. 12 (Mon. Boica 33, 11). – Vgl. GEBELE, Hermann, 16–17; KÄUPER, Verdun, Konstanz und Augsburg, 306–307; ZOEPFL, Augsburger Bischöfe im Investiturstreit, 317–318; ZOEPFL, Bistum Augsburg, 110–111.

<sup>170</sup> 1099 April 30. Intervenient in einer Urkunde für das Kloster Kremsmünster (DH.IV. 625 n. 463): *digno interventu [...] Herimanni Augustensis electi*. – Vgl. BÖHMER-LUBICH [RI III,2,3 (4)] 117 n. 1436; DIESTELKAMP-ROTTER, Urkundenregesten 1, 101 n. 147; MEYER VON KNONAU, Jbb. 5, 62–63; SCHMID, Regensburg [Königspfalzen 5,1.2] 147 n. 149; VOLKERT-ZOEPFL, Regesten 1, 230 n. 370.

<sup>171</sup> (1099 Herbst). – Vgl. MEYER VON KNONAU, Jbb. 5, 99; VOLKERT-ZOEPFL, Regesten 1, 232 n. 373.

<sup>172</sup> (1100) April 7. Begründung für die lange ausgebliebene Antwort und Freude, dass Hermann zur kirchlichen Einheit zurückgekehrt sei. Über die Rückkehr Hermanns in den Schoß der Kirche hatte den Papst im Jahr 1100 Bischof Gebhard III. von Konstanz informiert. – Vgl. BRACKMANN, Augustensis [Germ. Pont. II,1] 33 n. 19; JL 5825; LADEWIG, Regesten, 74 n. 591; MEYER VON KNONAU, Jbb. 5, 100; PAULUS, Bischof Hermann von Augsburg, 92–93; VOLKERT-ZOEPFL, Regesten 1, 233 n. 375, 235 n. 378; ZOEPFL, Augsburger Bischöfe im Investiturstreit, 319.

<sup>173</sup> (1100–1102). Bezüglich einer Angelegenheit im Augsburger Domkapitel. – Vgl. BRACKMANN, Augustensis [Germ. Pont. II,1] 34 n. 20, 47 n. 1; JL 5887; MEYER VON KNONAU, Jbb. 5, 146 Anm. 61; VOLKERT–ZOEPLF, Regesten 1, 234 n. 377.

<sup>174</sup> 1102 Dezember 25. – Vgl. BÖHMER–LUBICH [RI III,2,3 (4)] 145 n. 1482; ZOEPLF, Bistum Augsburg, 111–112.

<sup>175</sup> 1103 Ende Januar–Anfang Februar. – Vgl. MEYER VON KNONAU, Jbb. 5, 169, 177; VOLKERT–ZOEPLF, Regesten 1, 236 n. 381; ZOEPLF, Augsburger Bischöfe im Investiturstreit, 320; ZOEPLF, Bistum Augsburg, 111–112.

Eine umfangreiche Literaturliste findet sich immer im Anschluss an die Einzeleinträge der Bischöfe zu jedem Bistum.

Die digitale Transformation der Daten – insgesamt umfasste die als Word-Datei angelegte Bischofsliste zu Projektbeginn knapp 1100 Seiten, in der mehr als 5500 mit Anmerkungen belegte Handlungen verzeichnet waren – verlangte nach einer digitalen Aufbereitung, die Antworten zu geben versprach auf die an das Projekt gerichteten Forschungsfragen nach personalen Beziehungsgeflechten und die auch in der Lage war, komplexe Muster abzubilden. Aus diesen Erwägungen heraus fiel die Wahl auf die Graphdatenbank Neo4j.

### 3 Voraussetzungen für die digitale Transformation der Daten

Einen umfassenden Einblick in die Möglichkeiten und Funktionsweisen der Graphdatenbank Neo4j gibt Andreas KUCZERA in einer digital zur Verfügung gestellten und laufend aktualisierten Handreichung,<sup>12</sup> auf die auch im weiteren Verlauf mehrfach Bezug genommen wird. In Neo4j werden die Daten nicht in Form einer relationalen Aufbereitung (Tabellen bzw. SQL-Datenbank) strukturiert, sondern in Graphen mit Knoten und Kanten. Dies ermöglicht die Abbildung personeller Netzwerke in ihrem Zusammenspiel von gerichteten (also direkten) und transitiven (also indirekten) Verbindungen zueinander. Dadurch lassen sich Muster erkennen, die in klassisch relational aufgearbeiteten Datenbanken oft unentdeckt bleiben.<sup>13</sup> Graphen dienen damit der Modellierung und Analyse hochvernetzter Forschungsdaten und versprechen insbesondere durch ihre Möglichkeit, transitive personale

<sup>12</sup> Vgl. Kuczera, Graphentechnologien in den Digitalen Geisteswissenschaften. <https://kuczera.github.io/Graphentechnologien/> (Zugriff: 10.03.2025). Zur Anwendungsmöglichkeit von Graphentechnologie in den Geschichtswissenschaften, konkret in der Mediävistik, vgl. bereits Kuczera 2017 und Kuczera 2019.

<sup>13</sup> Vgl. Kuczera, Graphentechnologien in den Digitalen Geisteswissenschaften. <https://kuczera.github.io/Graphentechnologien/> (Zugriff: 10.03.2025).

Verbindungen aufzudecken, für das hier vorgestellte Projekt ein großer Gewinn zu sein. Die Grundlage bildet ein sogenannter *Labeled Property Graph*, ein Graph also, dessen Knoten und Kanten über Typen (*labels*) und Eigenschaften (*properties*) verfügen. Damit ist es möglich, die aus den Forschungsdaten zu extrahierenden Entitäten – in diesem Zusammenhang verstanden als für sich abgeschlossene Informationseinheiten, die eindeutig identifiziert werden können wie etwa Personen und Orte, aber auch jede Form eines Ereignisses (*events*) – abzubilden und miteinander in Bezug zu setzen (vgl. für ein allgemeines Schema Abb. 1).

Für den Import der Daten in die Graphdatenbank Neo4j sind jedoch Vorarbeiten zu erbringen, die entscheidend für die späteren Abfragemöglichkeiten innerhalb der Datenbank sind. Grundsätzlich ist es notwendig, die innerhalb der Word-Datei vorliegenden Informationen zum Import zunächst innerhalb einer Tabelle aufzubereiten. Die aus den Daten gewonnenen Aussagen sind in Form einer Ereignismodellierung angelegt. Jede Aussage besteht aus Subjekt, Prädikat und Objekt.<sup>14</sup> Diese Aufbereitung stellte – und stellt weiterhin – eine große Herausforderung dar und machte bereits mehrfach Anpassungen und Überarbeitungen notwendig. Die besondere Schwierigkeit ergibt sich aus der erforderlichen Aufgliederung und Vereinfachung der komplexen, ineinander verwobenen Aussagen zu den einzelnen Bischöfen in Form der genannten Ereignismodellierung mit klarer Subjekt-Prädikat-Objekt-Struktur. Der dafür gewählte Weg soll im Folgenden etwas detaillierter aufgezeigt werden. Als Beispiel dient weiterhin Bischof Hermann von Augsburg, den Ausgangspunkt bilden die in Kapitel 2 präsentierten Auszüge aus dem zu seiner Person und seinen personalen Beziehungen verfassten Artikel.

Die Überführung der Daten in tabellarische Form gelingt aktuell in Gestalt einer sukzessive wachsenden und den jeweils neu auftretenden Anforderungen entsprechend ausgestalteten Tabelle, die bei erfolgreicher Umsetzung innerhalb dieses Vorhabens auch bei anderen Projekten Anwendung finden soll (vgl. Abb. 2). Abgebildet und aufgeschlüsselt wird jede einzelne Handlung, an der entweder Bischof Hermann aktiv oder passiv beteiligt war oder die in seiner Bischofsstadt unter Beteiligung des Königs oder des Papstes stattgefunden hat. Jedes Ereignis (*event*) erhält einen eindeutigen *Universally Unique Identifier* (UUID), durch den das Ereignis direkt mit anderen in Bezug gesetzt werden kann. Die eigentliche Handlung setzt sich aus einem Subjekt, dem eigentlichen Ereignis und einem dazugehörigen Objekt zusammen. Nur in wenigen Fällen kommt die Handlung ohne Verknüpfung mit einem Objekt aus, etwa beim Tod des Subjektes. Schließlich

---

<sup>14</sup> Die Form der Überführung ähnelt sehr einer Aufbereitung in einem RDF (*Resource Description Framework*)-Datenmodell. Hierbei werden Aussagen als Triple modelliert, wobei jede Aussage gleichermaßen aus Subjekt, Prädikat und Objekt besteht. Mit der Modellierung der Bischofsnetzwerke in RDF-Form ist ebenfalls experimentiert worden, doch gestaltete sich die Umsetzung gerade komplexerer Ereignisse in RDF als nicht zufriedenstellend umsetzbar.



in der Haupttabelle Personen, die an verschiedenen Ereignissen beteiligt gewesen sind, über die Angabe der IDs eindeutig zu identifizieren und einander zuzuordnen.

Bischof Hermann von Augsburg hielt sich beispielsweise im Januar des Jahres 1104 gemeinsam mit Kaiser Heinrich IV. in Regensburg auf (vgl. Abb. 2, Zeile 279). Ersichtlich wird dies aus einer dort ausgestellten Urkunde des Kaisers über den Verzicht Hermanns auf von seinem Domkapitel zu Unrecht entzogene Güter.<sup>15</sup> Neben Hermann waren weitere Personen an der Urkundenausstellung als Aussteller (der Kaiser) sowie als Intervenienten (weitere Bischöfe) beteiligt, die entsprechend in der Tabelle unter mit *appearanceWikidata* und *appearance* bezeichneten Spalten vermerkt werden – je nachdem, ob sie bereits über eine ‚Wikidata‘-ID verfügen oder nicht. Da die weiteren Personen – mit einer Ausnahme – bereits über ihre entsprechenden IDs definiert wurden (es handelt sich um Kaiser Heinrich IV., die Erzbischöfe Friedrich I. von Köln und Bruno von Trier, den Gegenerzbischof Berthold von Salzburg sowie die Bischöfe von Bamberg, Eichstätt, Freising, Hildesheim, Minden, Münster, Osnabrück, Speyer und Würzburg), werden sie unter diesen IDs verzeichnet – entsprechend: Q60094 | Q77259 | Q1279091 | Q1597729 | Q1620545 | Q75915 | Q2567804 | Q1010354 | Q828126 | Q1697742 | Q993987 | Q1335275 | Q1355691. Die Interventionstätigkeit der genannten Bischöfe ist zusätzlich über die *relatedTo*-Spalte über eine UUID mit dem Ereignis selber, also der Urkundenausstellung in Regensburg, verknüpft. Damit sind automatisch sämtliche in Regensburg versammelten (Erz-)Bischöfe mit dem Ereignis in Verbindung zu bringen und in der Graphdatenbank entsprechend abfragbar. Sollten für eine Person noch keine Normdaten wie etwa eine ‚Wikidata‘-ID vergeben worden sein – zum Zeitpunkt der Niederschrift dieses Beitrages hatte von allen mit dem vorgenannten Ereignis in Verbindung stehenden Bischöfen einzig Bischof Adalbero von Trient noch keine ID –, wird der Name der Person in der *appearance*-Spalte aufgeführt und ohne Angabe einer ID innerhalb der Entitätentabelle definiert. Hinzugenommen wurden Spalten für unsichere Beteiligungen von Personen, auch hier geschieden zwischen denjenigen mit und ohne Normdaten. Weitere in die Tabelle integrierte Informationen betreffen die Datierung der einzelnen Ereignisse sowie Hinweise zu Quellen oder Literatur, aus denen die Informationen zum Ereignis geschöpft sind. Da viele Ereignisse über keine eindeutige Datierung verfügen und nur näherungsweise angegeben werden können, gibt es in der Tabelle zu jeder Datierung eine *notBefore*- und eine *notAfter*-Spalte. Ist ein Ereignis eindeutig einem Datum zuzuordnen – wie etwa die Beurkundung über den Verzicht Hermanns von Augsburg über die seinem Domkapitel zu Unrecht entzogenen Güter am 14. Januar 1104 –, sind die Angaben beider Datierungsspalten identisch. Schließlich wird, wie bereits oben kurz

---

<sup>15</sup> Vgl. Die Urkunden der deutschen Könige und Kaiser 483, S. 658 f., ausgestellt am 14. Januar 1104.

1	notBefore	notAfter	comment	source	sourceDetail	text	footnote
268	1096-12-01	1097-03-31	Udalschalk, De f	Udalscha ▾	c. 12	Investiert Dezember 1096	
269	1099-04-30	1099-12-31		Udalscha ▾	c. 12	Investiert Dezember 1096	
270	1133-03-19	1133-03-19		Heimo, C ▾	1133	Gestorben am 19. März 11	
271	1096-12-01	1099-12-31		Regester ▾	230 n. 371	Hermann war zum Zeitpun	
272	1096-12-01	1099-12-31		Regester ▾	230 n. 371	Hermann war zum Zeitpun	
273	1099-04-30	1099-12-31	ggf. auch schon	Udalscha ▾	c. 12	Seine Weihe empfang er vo	
274	1099-10-01	1099-12-31	mehrere Briefe	Regester ▾	232, n. 373	Ende des Jahres richtete H	
275	1100-04-07	1100-04-07	Antwort auf meh	Brackmar ▾	n. 19	auf die dieser im Folgejahr	
276	1100-04-07	1102-12-31		Brackmar ▾	n. 20, n. 1	sowie zwischen 1100 und 1	
277	1102-12-25	1102-12-25		RI III ▾	2,3 (4)	1102 feierte Hermann in M	
278	1103-01-20	1103-02-10	im Auftrag Heinr	Regester ▾	236, n. 381	und geleitete in seinem Auf	
279	1104-01-14	1104-01-14	Hermann entsag	Diplomat: ▾	658, n. 483	Mitte Januar 1104 war Her	

**Abb. 3 |** Die letzten Spalten zum Beginn des Episkopats Hermanns von Augsburg. Tabelle: Lukas Schuhr u. Matthias Weber.

ausgeführt, zu jedem Ereignis eine Quelle exemplarisch benannt, der die zuvor angegebenen Informationen entnommen worden sind. Zur Einsetzung und Weihe Bischof Hermanns von Augsburg (vgl. Abb. 3, Zeile 268 u. 269) ist etwa ein um 1120 entstandener Text Udalschalks von Augsburg angegeben,<sup>16</sup> während bei anderen Ereignissen auf Quellen bündelnde Werke wie die Regesten der Bischöfe von Augsburg<sup>17</sup> oder Bände der ‚Germania Pontificia‘<sup>18</sup> verwiesen wird (vgl. Abb. 3, Zeile 271 u. 275). Sollte es zu einem Ereignis mehr als einen Quellenbeleg geben, gehen diese zusätzlichen Informationen bei der digitalen Transformation nicht verloren, vielmehr werden auch die ursprünglichen Inhalte von Text und Anmerkungsapparat in Gänze in die Tabelle überführt. Dazu dienen die abschließenden zwei Spalten *text* und *footnote* (vgl. Abb. 3). Während der Text der Bischofsliste bereits in die Tabelle überführt werden konnte, steht dieser Schritt für die Anmerkungen noch bevor. Das Material liegt aber bereits in angepasster Form vor,<sup>19</sup> sodass auch die in den Anmerkungen getroffenen Formatierungen und Hervorhebungen bei der digitalen Transformation berücksichtigt werden können.

## 4 Die Daten in der Graphdatenbank – erste Ergebnisse

Die Transformation der innerhalb der Tabelle zusammengestellten Daten in die Graphdatenbank Neo4j erfolgt über ein Importskript in Cypher, der für die

<sup>16</sup> Vgl. Udalschalk: De Eginone et Herimanno.

<sup>17</sup> Vgl. Volkert u. Zoepfl 1955–1985.

<sup>18</sup> Vgl. Brackmann 1923.

<sup>19</sup> Gedankt sei an dieser Stelle explizit Hans-Werner Barz von der Digitalen Akademie der Akademie der Wissenschaften und der Literatur in Mainz.

Datenbank Neo4j erstellten Abfragesprache. Neo4j ist eine Open-Source-Software, die administrationsarm betrieben werden kann. Tutorials erleichtern den Einstieg zusätzlich. Zunächst wird in der Datenbank zu den Bischöfen der Knoten ‚Hermann von Augsburg‘ aufgerufen, der weiterhin als Beispiel dienen soll. Notwendig ist dafür folgende Abfrage, bei der bereits auf die hinterlegte ‚Wikidata‘-ID Bezug genommen wird:

```
| match (e:Entity {wikidataId:'Q61140624'})
| return *;
```

Gesucht wird nach der Entität mit der genannten ‚Wikidata‘-ID, die direkt zu Hermann von Augsburg führt (vgl. Abb. 4). In einem nächsten Schritt ist es möglich, sich sämtliche Ereignisse anzeigen zu lassen, mit denen die Entität ‚Hermann‘ mit der entsprechenden ‚Wikidata‘-ID auf Grundlage der Vorarbeiten in der Tabelle verknüpft ist. Momentan stehen 76 Ereignisse mit Hermann in Verbindung, Resultat ist ein schwer zu überblickender Graph (vgl. Abb. 5) – wengleich er etwa durch unterschiedliche Farbgebung oder Kantenbreite übersichtlicher gestaltet werden kann. In diesem Fall gilt es aber, die Abfrage zu konkretisieren und auf einen Beispielkomplex auszurichten. Ein Ziel des Projektes ist, die Abbildung der Verhältnisse zwischen den Bischöfen auf der einen sowie zwischen den Königen und Kaisern beziehungsweise den Päpsten auf der anderen Seite zu ermöglichen. Derartige personale Verbindungen zu ermitteln, um Personenkonstellationen in einem – zum Beispiel – zeitlichen, räumlichen oder familiären Rahmen nachgehen zu können, ist äußerst zeitaufwendig. Entsprechend erleichtert die Überführung großer Datenbestände in Graphdatenbanken die Durchsuchbarkeit und Visualisierbarkeit. Hermann stand mit beiden Parteien – König und Papst – in engerem, wengleich teils sehr angespanntem Verhältnis. Dies lässt sich gut anhand der in den Jahren 1106 und 1107 ausgerichteten päpstlichen Synoden von Guastalla (Norditalien, östlich von Parma) und Troyes (gelegen etwa 140 Kilometer südöstlich von Paris) verdeutlichen, deren Repräsentation in der Graphdatenbank gleichzeitig deren Möglichkeiten veranschaulicht. Zum Zeitpunkt der Synode von Guastalla im Oktober 1106 war König Heinrich V. noch kein ganzes Jahr im Amt. Nach der Entmachtung seines Vaters Heinrich IV. war er erst im Januar 1106 zum König erhoben worden, der militärische Widerstand seines Vaters war erst mit dessen Tod im August 1106 erloschen.<sup>20</sup> Im Rahmen der Synode von Guastalla sollte nun die seit langem herrschende Entfremdung zwischen Papst und deutschem König überwunden werden, weshalb Heinrich V. eine größere

---

<sup>20</sup> Vgl. exemplarisch Weinfurter 1992.

Node properties	
<b>&lt;elementId&gt;</b>	4:41a99bc0-abb0-4033-85c4-00662c8bf0e5:1293
<b>&lt;id&gt;</b>	1293
<b>label</b>	Hermann (Augsburg)
<b>type</b>	Person
<b>wikidataid</b>	Q61140624
<b>wikidataUrl</b>	https://www.wikidata.org/wiki/Q61140624

Abb. 4 | Die Repräsentation Hermanns von Augsburg im Graph. Graphik: Matthias Weber.

Overview

**Node labels**

- Entity (11)
- Event (76)

**Relationship types**

- IS\_SUBJECT\_IN (42)
- IS\_OBJECT\_IN (20)
- APPEARS\_IN (22)
- RELATED\_TO (29)

Displaying 77 nodes, 112 relationships.

Abb. 5 | Die Entität Hermann (lila) im Graph inmitten der mit ihm verknüpften Ereignisse (events, orange). Graphik: Matthias Weber.

Gesandtschaft zu dieser Synode auf den Weg brachte.<sup>21</sup> Hermann von Augsburg war nicht Teil dieser Delegation, aber dennoch in Guastalla anwesend. Ihm drohte aufgrund seiner simonistischen Investitur durch Kaiser Heinrich IV. die Absetzung. Schlussendlich wurde er suspendiert und das endgültige Urteil auf einen späteren Zeitpunkt vertagt.<sup>22</sup> Wenige Monate später fand im Mai 1107 eine weitere päpstliche Synode in Troyes statt.<sup>23</sup> Die Beziehungen zwischen Papst und deutschem König hatten sich in den zurückliegenden Monaten sichtlich verschlechtert. In der Folge hatten sich keine deutschen Bischöfe in Troyes eingefunden,<sup>24</sup> ein Großteil der Abwesenden, darunter auch Bischof Hermann von Augsburg, wurde daraufhin von Papst Paschalis II. mit der Suspension belegt. Die Graphdatenbank ermöglicht es nun, die beiden Synoden in Beziehung zueinander zu setzen und personelle Verknüpfungen offenzulegen. Dazu können die Synoden einzeln aufgerufen werden (vgl. zur Synode von Guastalla Abb. 6), über eine etwas komplexere Abfrage können sie aber auch direkt miteinander in Verbindung gebracht werden (vgl. Abb. 7):

```

match (e:Event)-[:RELATED_TO*..2]-(s:Event)
  where e.subject = 'Synode'
  and e.notBefore in ['1107-05-23','1106-10-22']
  optional match (p:Entity)--(e), (s)--(p2:Entity)
return *
```

Zwei Ereignisse mit dem Subjekt ‚Synode‘ werden hiermit abgefragt, beide sind eindeutig über die Datumsangaben definiert, an denen sie stattgefunden haben. Zusätzlich wird nach den mit beiden Synoden verbundenen Ereignissen sowie allen damit in Beziehung stehenden Personen gefragt. Dadurch entsteht ein Beziehungsnetz, das zwar Papst Paschalis II. in zentraler Position zeigt – da er in der Graphdatenbank als Subjekt vieler Ereignisse in Zusammenhang mit den Synoden geführt wird –, aber dennoch nur als eine von vielen Personen, die an den Synoden von Guastalla und Troyes teilgenommen beziehungsweise gerade nicht teilgenommen haben, was sich in zahlreichen in Troyes ausgesprochenen Suspensionen niederschlug. Im Ergebnis entsteht ein dichtes Geflecht von Ereignissen und damit verbundenen Personen. Dabei werden den beiden Synoden zusätzliche Ereignisse zugeordnet, wodurch räumliche und zeitliche Übereinstimmungen und Zusammenhänge sichtbar gemacht werden können. Das hier graphisch vorgelegte Geflecht bildet zunächst nur einen Ausschnitt der zukünftig bestehenden Möglichkeiten ab.

21 Zur Synode von Guastalla vgl. Gresser 2006, S. 368–378.

22 Vgl. Lubich u. Weber [im Druck], n. 121.

23 Zur Synode von Troyes vgl. Gresser 2006, S. 378–385.

24 Abgesehen möglicherweise vom Salzburger Erzbischof Konrad; vgl. den Kommentar zu Lubich u. Weber [im Druck], n. 150.



Für seine Vervollständigung ist die Überführung sämtlicher Bischöfe und Bistümer des Untersuchungszeitraums in eine tabellarische Form nötig, was dank der bisher geleisteten Vorarbeiten nun möglich ist. Dadurch können nicht nur die Verbindungen von Bischöfen zu König/Kaiser und Papst visualisiert werden, sondern etwa auch Rückschlüsse gezogen werden auf die bischöflichen Netzwerke, zu denen die Quellen häufig keine oder nur implizite Andeutungen und Hinweise geben. Die Visualisierung der Daten kann dabei zu neuen Erkenntnissen beitragen.

## 5 Fazit und Ausblick

Es sollte gezeigt werden, dass die Überführung der Forschungsdaten zu den Bischöfen des deutschen Reiches nördlich der Alpen in einen Graphen zu einer explorativen Auseinandersetzung mit diesen Daten führen kann, die nicht chronologisch oder linear bestimmt ist und ganz unterschiedlichen Fragestellungen offensteht. Die Visualisierung im Graph bildet dabei kein endgültiges Ergebnis ab, sondern fügt sich ein in den Prozess der Beantwortung anfangs gestellter Fragen.

Das hier knapp skizzierte Verfahren zur Transformation großer Datenbestände in eine graphbasierte Umgebung – vorgestellt am Beispiel eines Datenbestandes zu den nordalpinen Bischöfen in Deutschland im 11. und beginnenden 12. Jahrhundert – steht ungeachtet aller Vorarbeiten noch am Anfang. Insgesamt zeigt sich aber schon in diesem frühen Stadium, dass die Überführung der Forschungsdaten in eine Graphmodellierung die Möglichkeit bietet, diese Daten in neuen Zusammenhängen zu erschließen, verschiedenste Forschungsfragen an diese zu adressieren und neue Abfragen innerhalb dieser Daten durchzuführen. Die Überführung der einzelnen Ereignisse in eine tabellarische Form wird weiterhin Anpassungen und Modifikationen nötig machen, damit die einzelnen bischöflichen Handlungen bestmöglich wiedergegeben werden können und langfristig auch die Verknüpfung der Graphdatenbank mit weiteren Datenbanken gewährleistet ist. Die Abfrage innerhalb der Graphdatenbank erfordert aktuell noch Kenntnisse der Sprache Cypher, hier wird über die Implementierung benutzerfreundlicher Interfaces nachgedacht, um die Verwendung zu erleichtern. Grundsätzlich wird die Modellierung kontinuierlich verbessert, um das Projekt für viele künftige Forschungsprojekte und Anwendungsfälle in den *Digital Humanities* weiterzuentwickeln.<sup>25</sup>

---

<sup>25</sup> Ein besonderer Dank gilt Prof. Dr. Andreas Kuczera (Technische Hochschule Mittelhessen, Gießen) für die intensive Begleitung und Förderung des Projektes sowie für seine unschätzbare digitale Expertise, ohne die dieses Projekt die Transformation in eine graphbasierte Umgebung niemals erlebt hätte, und Lukas Schuhr (Ruhr-Universität Bochum) für seine unermüdliche Arbeit am Projekt, wodurch überhaupt erst die Voraussetzungen gegeben waren, eine digitale Transformation ins Auge zu fassen.

## Literaturverzeichnis

## Quellen

- Brackmann, Albert (Bearb.):** *Germania pontificia sive repertorium privilegiorum et litterarum a Romanis pontificibus ante annum MCLXXXVIII Germaniae ecclesiis monasteriis civitatibus singulisque personis concessorum*. Bd. 2,1: *Provincia Maguntinensis 1: Dioeceses Eichstetensis, Augustensis, Constantiensis I*. Berlin 1923.
- Engels, Odilo u. Stefan Weinfurter (Hgg.):** *Series episcoporum ecclesiae catholicae occidentalis ab initio usque ad annum MCXCVIII*. Series V: *Germania*. Tomus I: *Archiepiscopatus Coloniensis*. Stuttgart 1982.
- Engels, Odilo u. Stefan Weinfurter (Hgg.):** *Series episcoporum ecclesiae catholicae occidentalis ab initio usque ad annum MCXCVIII*. Series V: *Germania*. Tomus II: *Archiepiscopatus Hammaburgensis sive Bremensis*. Stuttgart 1984.
- Kluger, Helmuth (Hg.):** *Series episcoporum ecclesiae catholicae occidentalis ab initio usque ad annum MCXCVIII*. Series VI: *Britannia, Scotia et Hibernia, Scandinavia*. Tomus II: *Archiepiscopatus Lundensis*. Stuttgart 1992.
- Lubich, Gerhard (Bearb.):** *Die Regesten des Kaiserreiches unter Heinrich IV. 1056 (1050)–1106*. 3. Lief.: 1076–1085. Nach Vorarbeiten von Tilman Struve unter Mitwirkung von Dirk Jäckel. Köln, Weimar, Wien 2016a.
- Lubich, Gerhard (Bearb.):** *Die Regesten des Kaiserreiches unter Heinrich IV. 1056 (1050)–1106*. 4. Lief.: 1086–1105/06. Nach Vorarbeiten von Daniel Brauch unter Mitwirkung von Matthias Weber. Köln, Weimar, Wien 2016b.
- Lubich, Gerhard (Bearb.):** *Die Regesten des Kaiserreiches unter Heinrich IV. 1056 (1050)–1106*. 5. Lief.: *Die Regesten Rudolfs von Rheinfelden, Hermanns von Salm und Konrads (III.) Verzeichnisse, Register, Addenda und Corrigenda*. Unter Mitwirkung von Dirk Jäckel, Cathrin Junker, Markus Keller, Lisa Klocke u. Matthias Weber. Köln, Weimar, Wien 2018.
- Lubich, Gerhard u. Matthias Weber (Bearb.):** *Die Regesten des Kaiserreiches unter Heinrich V. 1106 (1086)–1125*. 1. Lief.: 1086–1111. Bearb. v. Gerhard Lubich u. Matthias Weber. Köln, Weimar, Wien [im Druck].
- Udalschalk:** *De Eginone et Herimanno*. Hrsg. v. Philipp Jaffé (MGH SS 12). Hannover 1851, S. 429–447.
- Die Urkunden der deutschen Könige und Kaiser. Bd. 6:** *Die Urkunden Heinrichs IV.* Hrsg. v. Dietrich von Gladiß u. Alfred Gawlik (MGH DD 6). Hannover 1951–1978.
- Volkert, Wilhelm u. Friedrich Zoepfl (Bearb.):** *Die Regesten der Bischöfe und des Domkapitels von Augsburg*. Bd. 1 (Veröffentlichungen der Schwäbischen Forschungsgemeinschaft 2b,1). Augsburg 1955–1985.
- Watt, Donald (Hg.):** *Series episcoporum ecclesiae catholicae occidentalis ab initio usque ad annum MCXCVIII*. Series VI: *Britannia, Scotia et Hibernia, Scandinavia*. Tomus I: *Ecclesia Scoticana*. Stuttgart 1991.

## Forschungsliteratur

**Bihrer, Andreas u. Stephan Bruhn (Hgg.):**

Jenseits des Königshofs. Bischöfe und ihre Diözesen im nachkarolingischen ostfränkisch-deutschen Reich (850–1100) (Studien zur Germania Sacra N. F. 10). Berlin, Boston 2019.

**Bihrer, Andreas u. Hedwig Röckelein**

(Hgg.): Die ‚Episkopalisierung der Kirche‘ im europäischen Vergleich / The ‚Episcopalization of the Church‘ in European Comparison (Studien zur Germania Sacra N. F. 13). Berlin, Boston 2022.

**Bode, Tina:** König und Bischof in ottonischer Zeit: Herrschaftspraxis, Handlungsspielräume, Interaktionen (Historische Studien 506). Husum 2015.

**Gresser, Georg:** Die Synoden und Konzilien in der Zeit des Reformpapsttums in Deutschland und Italien von Leo IX. bis Calixt II. 1049–1123 (Konziliengeschichte. Reihe A. Darstellungen). Paderborn u. a. 2006.

**Johrendt, Jochen:** Der Investiturstreit (Geschichte kompakt). Darmstadt 2018.

**Kuczera, Andreas:** Graphentechnologien in den Digitalen Geisteswissenschaften. Modellierung – Import – Analyse. <https://kuczera.github.io/Graphentechnologien/> (Zugriff: 10.03.2024).

**Kuczera, Andreas:** Graphentechnologien in den Digitalen Geisteswissenschaften. In: ABI Technik 37 (2017), S. 179–196.

**Kuczera, Andreas:** Die ‚Regesta Imperii‘ im digitalen Zeitalter. Das Regest als Netzwerk von Entitäten. In: Roman Bleier, Franz Fischer, Torsten Hiltmann u. a. (Hgg.): Digitale Mediävistik (Das Mittelalter. Perspektiven mediävistischer Forschung 24). Berlin 2019, S. 157–172.

**Ott, John S. u. Anna Trumbore Jones:**

Introduction: The Bishop Reformed. In:

Dies. (Hgg.): The Bishop Reformed. Studies of Episcopal Power and Culture in the Central Middle Ages (Church, Faith, and Culture in the Medieval West). Aldershot, Burlington 2007, S. 1–20.

**Patzold, Steffen:** Episcopus: Wissen über Bischöfe im Frankenreich des späten 8. bis frühen 10. Jahrhunderts (Mittelalter-Forschungen 25). Ostfildern 2008.

**Scheibelreiter, Georg:** Der Bischof in merowingischer Zeit (Veröffentlichungen des Instituts für österreichische Geschichtsforschung 27). Wien, Köln, Graz 1983.

**Weinfurter, Stefan:** ‚Series episcoporum‘ – Probleme und Möglichkeiten einer Prosopographie des früh- und hochmittelalterlichen Episkopats. In: Neithard Bulst u. Jean-Philippe Genet (Hgg.): Medieval Lives and the Historian. Studies in Medieval Prosopography. Proceedings of the First International Interdisciplinary Conference on Medieval Prosopography, University of Bielefeld, 3–5 Dezember 1982. Kalamazoo 1986, S. 97–112.

**Weinfurter, Stefan:** Reformidee und Königtum im spätsalischen Reich. Überlegungen zu einer Neubewertung Kaiser Heinrichs V. In: Ders. (Hg.): Reformidee und Reformpolitik im spätsalisch-frühstaufischen Reich (Quellen und Abhandlungen zur mittelhochrheinischen Kirchengeschichte 68). Mainz 1992, S. 1–45.

**Zey, Claudia:** Der Investiturstreit. München 2017.

**Zielinski, Herbert:** Der Reichsepiskopat in spätottonischer und salischer Zeit (1002–1135). Stuttgart 1984.

# Zeit und Raum

## Visualisierung und Analyse von Geodaten über Transaktionen mit Basler Liegenschaften in der Vormoderne

### Kontakt

**PD Dr. Benjamin Hitz,**

Universität Basel, Departement  
Geschichte, Hirschgässlein 21,  
4051 Basel, Schweiz,  
benjamin.hitz@unibas.ch

 <https://orcid.org/0000-0002-3208-4881>

**Prof. Dr. Tobias Hodel,** Universität

Bern, Walter Benjamin Kolleg,  
Digital Humanities,  
Muesmattstrasse 45,  
3012 Bern, Schweiz,  
tobias.hodel@unibe.ch

 <https://orcid.org/0000-0002-2071-6407>

**Abstract** This paper explores how information extraction and geographic information systems (GIS) can be combined to open new perspectives for urban economic history research. Using the example of Basel's old town between 1300 and 1700, we analyse 'economic activities related to houses' and demonstrate the entire process from data acquisition and preparation to visualisation. Drawing on the 'Historisches Grundbuch Basel' (Historical Land Registry of Basel), textual data is linked to geographical information to identify spatiotemporal patterns. Automatic text recognition, georeferencing, and visualisation algorithms (in particular, heat maps, hexagonal grids, and cluster analyses) are employed for data preparation and analysis. These methods capture the city's economic and social dynamics and enable the formulation of new research questions. The results highlight the importance of careful data preparation and the marking of uncertainties as well as the role of algorithmic applications in historical research. Our work underlines that visualisations do not simply present final results; instead, they serve as intermediate forms of analysis that open up new possibilities for insight, question existing assumptions, and foster serendipitous discoveries.

**Keywords** Data Preparation; Digital Medieval Studies; Geovisualisation; Information Extraction; Machine Learning

## 1 Einleitung

Die Wirtschaftsgeschichte der Vormoderne beschäftigt sich seit längerer Zeit mit der Entwicklung von Konzepten wie Besitz und Besitzübertragung, Herrschaft und Kommerzialisierung. Eine besondere Bedeutung kommt dabei der Frage zu, wie mit unterschiedlichen Formen von Kapital und Besitz und unterschiedlichen Formen ihrer Bewirtschaftung umzugehen ist, wie sie sich insbesondere am Ende dieser Epoche manifestieren.<sup>1</sup> Um diesen Phänomenen beizukommen, bedarf es für die Vormoderneforschung<sup>2</sup> besonderer Anstrengungen, da die Überlieferungssituation komplex und unvollständig ist. Dass die Transaktionen in ihrer sozialen Einbettung häufig in vielfältiger Weise überlagert waren, erschwert ihre Nachvollziehbarkeit zusätzlich.<sup>3</sup>

Unser Beitrag greift diese Problematik auf und demonstriert den Prozess von der Datengewinnung bis zur Visualisierung der Ergebnisse im Zuge der Analyse des ‚Wirtschaftens mit Häusern‘ in Basel. Dabei steht die Reflexion des Umgangs mit digitalen Daten im Vordergrund, seien es digitale Methoden der Datenerhebung und -verarbeitung, seien es Formen der Aggregation und Visualisierung von Daten. Ziel ist es, neue Perspektiven für die Erforschung der Wirtschaftsgeschichte im Mittelalter zu eröffnen und zu diskutieren sowie die Bedeutung digitaler Werkzeuge im historischen Forschungsprozess zu unterstreichen. Als leitende Fragestellung ist denn auch zu eruieren, wie Informationen aus komplexen und quellenkritisch umsichtig zu behandelnden Dokumenten extrahiert und vor allem auch in geographische Visualisierungen integriert werden können. Kürzer gefasst lautet also die Frage: Lässt sich eine dichte Datensammlung zu häuserbezogenen Transaktionen einer Stadt auch für einen mediävistischen Erkenntnisgewinn nutzen? Und welche Rolle spielen dabei Geovisualisierungen?

---

1 Der Beitrag geht auf die Überlegungen zur Umsetzung von Daten zurück, die im Rahmen des Projekts ‚Ökonomien des Raums. Praktiken, Diskurse und Akteure auf dem Basler Immobilienmarkt (1400–1700)‘ erarbeitet wurden. Neben den beiden Autoren beteiligen sich Lucas Burkart, Jonas Aebi, Katrin Fuchs, Ismail Prada und Aline Vonwiller am Projekt, das durch den Schweizerischen Nationalfonds, <https://data.snf.ch/grants/grant/208093> (Zugriff: 03.12.2024) gefördert wird. Für die gute Zusammenarbeit möchten wir uns beim Staatsarchiv Basel-Stadt, dem Grundbuch- und Vermessungsamt Basel-Stadt, insbesondere Andreas Kettner, und der Denkmalpflege Basel-Stadt bedanken. Daten und Dokumentation werden regelmäßig aktualisiert und auf GitHub publiziert.

2 Dem epochalen Fokus dieser Publikation entsprechend, werden wir den Fokus auf die Zeit vor 1500 legen. Da aber in der Wirtschaftsgeschichte die Epochengrenze keine deutliche Zäsur mit sich brachte, untersuchen wir auch Quellenbelege aus dem 16. Jahrhundert und sprechen insgesamt vom vormodernen Wirtschaften mit Liegenschaften.

3 Zur Wirtschaftsgeschichte des Spätmittelalters vgl. einführend Gilomen 2014. Zu Handel und Austauschformen vgl. Howell 2010. Spezifischer zur Thematik des vormodernen Immobilienmarktes vgl. van Bavel 2011; Lorenzen-Schmidt 2014; Hammel 1990.

Eine datengetriebene Auswertung stützt sich notwendigerweise auf eine Zusammenarbeit mit Institutionen wie Archiven und Geodatenzentren, die Vorarbeiten zur Datenerhebung leisten und etwa historische Parzellenpläne zur Verfügung stellen. Für den vorliegenden Beitrag steht entsprechend auch nicht die Datenvernetzung über *Linked Data*-Ansätze im Zentrum, sondern die Datenaggregation aus unterschiedlichen Quellen. Als Ausgangspunkt stützen wir uns auf das ‚Historische Grundbuch Basel‘ (HGB), das bereits selbst als Zusammenführung aus Datenpunkten verstanden werden kann und weiter unten quellenkritisch eingeführt wird.

Unsere Fragestellung konzentriert sich darauf, wie umfangreiche, geographisch zuordenbare Daten in Kombination mit geographischen Gegebenheiten visualisiert werden können. Welche Methoden der Auswertung bieten sich an? Und welche Bedeutung kommt solchen Visualisierungen im Forschungsprozess zu? Diese Fragen führen uns zu einem umfassenden Forschungsansatz, der historische Daten im Raum verortet und auf digitale Methoden setzt. In diesem Beitrag zeigen wir diesen Ansatz anhand ökonomischer Daten zu Häusern der Basler Altstadt zwischen 1300 und 1700.

Die Überführung von Daten historischer Gebäude in georeferenzierte Daten ist komplex und mit Unsicherheiten behaftet. Er gelingt nur durch die Verknüpfung von Metadaten aus Archiven mit den Geodaten der städtischen Vermessung. Daher sind die sorgfältige Dokumentation der Datenentstehung und die Markierung der damit verbundenen Unsicherheiten entscheidend. Unser Quellenkorpus besteht aus Registerkarten, die jeweils eine Transaktion oder andere juristische Handlungen dokumentieren. Die Anreicherung dieser Textdaten mit geographisch-temporalen Informationen bildet einen zentralen Schritt unserer Datenaufbereitung, die Visualisierungen in Zeit und Raum anstreben. Die Visualisierung wiederum beeinflusst die Analyse, indem sie einerseits die Analyse selbst zentral unterstützt und andererseits neue Fragestellungen aufwirft und eine Thesenbildung ermöglicht. Anhand solcher Visualisierungen, im Beitrag demonstriert am Beispiel von Berufsbezeichnungen, Beschlagnahmungsverfahren und erstmaligen Erwähnungen von Häusern, die im HGB zwischen 1350 und 1550 dokumentiert sind, zeigen wir mögliche Wege der Datenexploration. Von einfachen Punktsammlungen über die Aggregation in Planquadraten und Dichteberechnungen bis hin zur Clusteranalyse können die gleichen Daten auf unterschiedliche Weise dargestellt, analysiert und interpretiert werden.

Wir referieren in diesem Beitrag zunächst den Forschungsstand mit Fokus auf Techniken der Datenerstellung und -visualisierung mit Geoinformationssystemen, um anschließend das Findmittel des HGB, das eine enorme Vielfalt an Informationen bündelt, als Datengrundlage für die (Klein-)Stadt einzuführen. Weiter zeigen und erläutern wir, welche Arbeitsschritte zur Überführung der wichtigsten Informationen aus textuellen Quellen- oder besser Exzerptsammlungen in eine

Datenbankform notwendig sind, damit geographische Auswertungsvisualisierungen generiert werden können. Schließlich testen wir unterschiedliche Aggregierungs- und Visualisierungsmuster, mit dem Ziel, an ihnen Möglichkeiten der Interpretation zeitlicher und geographischer Makroauswertungen zu erkunden.

## 2 Forschungsstand

### 2.1 Städtische Wirtschaftsgeschichte

In der Wirtschaftsgeschichte wird die Bedeutung von Städten als Wirtschafts- und Handelszentren mit ihrer Wirkung nach außen betont, in Bezug auf Immobilien gilt das Interesse der Forschung oft urbanen Investitionen in ruralen Grundbesitz.<sup>4</sup> Das vormoderne Wirtschaften mit städtischem Grund und Boden ist dagegen noch weit weniger untersucht und mit einer Vielzahl an heuristischen Hindernissen verbunden. Traditionellerweise fokussieren Arbeiten auf einzelne Quellengattungen, die für einen verhältnismäßig kurzen Zeitraum ausgewertet und deren Ergebnisse auf Teile des Stadtraums umgelegt werden, etwa Quellen zu Liegenschaftsverkäufen.<sup>5</sup> Oder aber es werden Personengruppen über einen längeren Zeitraum verfolgt.<sup>6</sup> Eine Gesamtsicht über die Bedeutung von Immobilien in einer städtischen Wirtschaft steht noch aus.

### 2.2 Geovisualisierungen

Datenvisualisierungen werden in der Forschung häufig auf umfangreiche geographische Entitäten angewendet und vor allem zur Vermittlung komplexer Informationen genutzt. In der Mediävistik wird ihr Einsatz vermehrt propagiert, was sich etwa daran zeigt, dass sie in jüngeren Einführungen zur historischen Arbeitstechnik erwähnt und in der Lehre verwendet werden.<sup>7</sup> Auch für Datensätze, die aus vormodernen Quellen gewonnen wurden, bieten sich Visualisierungen an, wie etwa von Pilgerrouen oder anderen Reisen.<sup>8</sup> Dabei wird gerne betont, dass sowohl umfangreichere Datenmengen als auch komplex modellierte Daten visuell

---

4 Vgl. Fouquet u. Andermann 2016; Krauer 2018.

5 Vgl. Ertl 2020.

6 Vgl. als Übersicht für Großbritannien Harkes 2022.

7 Für die Mediävistik vgl. Birnbaum, Bonde u. Kestemont 2017. Mit Blick auf die *Digital Humanities* schon sehr früh und allgemeiner zur Nutzung von Visualisierung Jessop 2008; Owens 2007.

8 Vgl. Perrone u. Traynor 2021; Sidiropoulos 2010.

verarbeitet werden können, beispielsweise um die Reisezeit nachvollziehbar zu machen.<sup>9</sup>

Im Umfeld der städtischen Wirtschaftsgeschichte kommen solche Visualisierungen mit Hilfe von Geoinformationssystemen (GIS) bisher jedoch nur vereinzelt zum Einsatz, wie beispielsweise in den Arbeiten von Roland GERBER über die Stadt Bern im Spätmittelalter, die den Fokus auf Steuerlisten gelegt haben.<sup>10</sup> Seine Auswertungen stützen sich hauptsächlich auf eine einzige Quellengattung. Die gewonnenen Visualisierungen ermöglichen Einsichten in die Sozialtopographie ökonomischer Eliten in einer Stadt wie Bern.

Diskussionen zum Design und der Anwendung von Visualisierungen sind im Kontext von geographischen Anwendungen für historische Daten selten anzutreffen. Die Überlegungen etwa von Johanna DRUCKER wurden bislang nur bedingt reflektiert und eigentlich nicht umgesetzt. So weist DRUCKER zu Recht darauf hin, dass geographische Darstellungen häufig problematische Skalen und unterkomplexe Formen der Visualisierung nutzen.<sup>11</sup> Sie stört sich insbesondere daran, dass die genutzten Daten eher simplizistisch aggregiert werden.

### 2.3 Nutzung von GIS und entsprechender Software

GIS hat sich zwar mittlerweile als Methode in der historischen Forschung durchgesetzt, wird jedoch häufig in einfachen Anwendungen genutzt, wie – um ein populäres Beispiel zu nennen – zur Darstellung von Herkunfts- und Wohnorten von Studenten in der Datensammlung des ‚Repertorium Academicum Germanicum‘.<sup>12</sup> Standardsoftware wie QGIS<sup>13</sup> und ArcGIS spielen bislang nur bedingt eine Rolle, obwohl sie auch komplexere räumliche Auswertungen erlauben würden.

Um historische Gegebenheiten adäquat mit heute gebräuchlichen Koordinatensystemen zu kombinieren, stehen häufig historische Karten und urbane Katasterpläne aus dem 19. Jahrhundert im Fokus, die selbst den Stand des 19. Jahrhunderts spiegeln. Mittels Kombination von historischer Darstellung und georeferenzierten Stadtkarten des 21. Jahrhunderts können historische Häuser und Parzellen zumindest näherungsweise kartiert werden.<sup>14</sup>

---

9 Vgl. für eine systematische Übersicht Perrone, Franklin-Lyons, Shaw u. a. 2022. Einführend ist noch immer Tufté 2001 leitend. Für eine allgemeine Einführung vgl. auch Heidmann 2013.

10 Vgl. insbesondere die Auswertungen in Gerber 2002.

11 Drucker 2011 und allgemeiner für die *Digital Humanities* Drucker 2015.

12 Vgl. Schwinges 2015.

13 QGIS Association 2024.

14 Zur Vorgehensweise vgl. z. B. Vetch, Clarke u. Lilley 2012.

## 2.4 ‚Historisches Grundbuch Basel‘: den ‚Datenschatz‘ heben

Das ‚Historische Grundbuch Basel‘ ist eine einmalige Informationssammlung zur Entwicklung der Stadt Basel und ihres Immobilienmarktes seit dem Aufkommen einer häuserbezogenen Überlieferung im 13. Jahrhundert bis ins Jahr 1800. Das Korpus wurde zwischen 1895 und 1935 im Staatsarchiv Basel-Stadt erarbeitet und besteht aus etwa 120.000 handschriftlichen Registerkarten. Es ist das Ergebnis einer systematischen Durchforstung der Basler Quellenbestände der Vormoderne. Quellenfunde wurden in Form von wörtlichen oder leicht modernisierten knappen Exzerpten wiedergegeben, die den für den Umgang mit Besitz und Rechten relevanten Teil aus den Quellen (wie etwa Urkunden, Zinsurbaren und Gerichtsakten) extrahieren. Diese Belege wurden in Dossiers gruppiert, die je ein einzelnes Haus der Basler Altstadt umfassen und nach einem Adressverzeichnis von 1862 strukturiert sind. Auch wenn die Exzerpte zum Teil aus dem Kontext gerissene und gekürzte Teile der Dokumente wiedergeben, erlauben sie eine tiefergehende Untersuchung der historischen Sprachverwendung (was hier jedoch nicht im Fokus steht). Da den Belegen jeweils Quellennachweise beigegeben sind, können die Exzerpte anhand der erhaltenen Quellenbestände überprüft werden. Der stichprobenartig unternommene Abgleich zwischen Exzerpt und Quelle zeigte eine sehr hohe Qualität und Genauigkeit der Exzerpte.

In unserem Projekt haben wir auf den vom Staatsarchiv Basel-Stadt gescannten Karteikarten des HGB mit Transkribus<sup>15</sup> eine *Handwritten Text Recognition* (HTR) und eine visuelle Segmentierung durchgeführt. Somit können wir auf die Transkripte der verschiedenen Textregionen zugreifen und die Inhalte der Karteikarten auswerten. Obwohl die Strukturierung der Informationen auf den Karteikarten den Interessen der Urheber:innen des HGB und ihren Konventionen folgt, bietet es mit seiner Masse an Informationen eine einzigartige Möglichkeit, die Entwicklung des städtischen Raums in Basel bis ins 19. Jahrhundert mit hoher Präzision nachzuvollziehen. Die Datenbasis ermöglicht Untersuchungen des vor-modernen Wirtschaftens mit städtischen Liegenschaften in synchroner und diachroner Perspektive sowie in räumlicher Skalierbarkeit von einzelnen Häusern bis zum gesamten Stadtraum. Damit bildet die Datengrundlage eine ausgezeichnete Basis für eine detaillierte Analyse und Reflexion der städtischen Entwicklungen und wirtschaftlichen Bedingungen im historischen Basel.

---

15 Zu Transkribus siehe <https://www.transkribus.eu/> (Zugriff: 03.12.2024) sowie Muehlberger, Seaward, Terras u. a. 2019.

### 3 Vom historischen Gebäude zum Punkt auf der Karte

Die oben bereits kurz beschriebene Aufbereitung der Textdaten aus dem HGB wurde in einem zweiten Schritt um Archivdaten ergänzt. Dabei profitieren wir nicht nur von der qualitativ hochwertigen Digitalisierung durch das Staatsarchiv Basel-Stadt, sondern auch von der Erschließung der Archivdaten, die bereits auf Stufe der Dossiers erfolgt. Damit stehen Metadaten bis auf die Stufe einzelner Adressen zur Verfügung.<sup>16</sup>

Das erstellte Datenmodell trennt zwischen den Metadaten und den extrahierten Informationen, aus denen logische Einheiten (Aussagen zu einem Vorgang, zum Beispiel zu einer Pfändung) erstellt wurden. Mit der im Projekt eingesetzten versionierten relationalen Datenbank ist es grundsätzlich möglich, automatisierte Schritte ebenso nachvollziehbar zu gestalten wie manuelle Zuordnungen, die in Einzelfällen vorgenommen wurden, etwa wenn die Text- oder Regionenerkennung ungenügend war.<sup>17</sup> Damit sind die Voraussetzungen geschaffen, um aus textuellen Daten Verknüpfungen mit Geodaten abzuleiten und für Auswertungen zu visualisieren.

#### 3.1 Häuserdossiers oder die Interpretation von Hausbeschreibungen

Für die Bearbeitenden des HGB war die genaue Identifikation von Häusern zugleich ein zentrales Anliegen und eine beachtliche Schwierigkeit. Anhand der in den Quellen üblichen Lagebeschreibungen von Häusern konnte in Kombination mit anderen Quellen die Nachbarschaft von Häusern ermittelt werden (Abb. 1). Aber wie sollten diejenigen Häuser beschrieben werden, die beim Erstellen des HGB zum Teil nicht mehr bestanden? Die Ersteller:innen wählten die Adressstruktur des späteren 19. Jahrhunderts, die zwar zum Zeitpunkt der Erarbeitung nicht mehr dem baulichen Zustand der Stadt entsprach, aber dank Stadt- und Katasterplänen genau bekannt war. Das Basler Adressbuch von 1862<sup>18</sup> entsprach weitgehend einem im gleichen Jahr erstellten Stadtplan, der nach seinem Erschaffer als Löffelplan bezeichnet wird.<sup>19</sup> Zudem besteht als weitere Informationsquelle

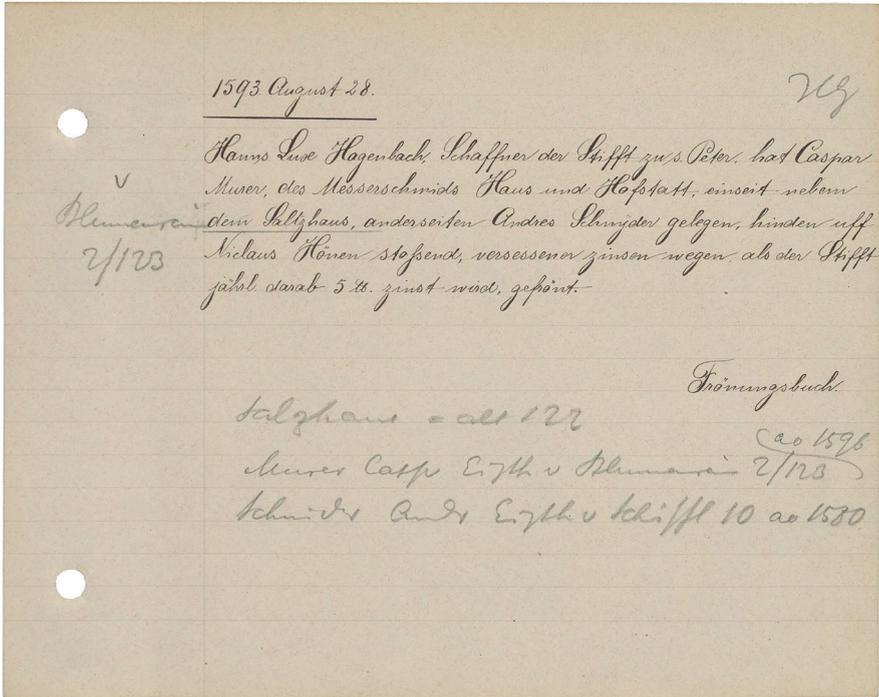
---

<sup>16</sup> Vgl. <https://dls.staatsarchiv.bs.ch/records/hierarchy?path=0001%2F0008%2F0002%2F0003%2F0001%2F0005&activeNode=1016781> (Zugriff: 03.12.2024).

<sup>17</sup> Zum Datenmodell siehe das *Repository* <https://github.com/history-unibas/Postgresql-Project-Database> (Zugriff: 03.12.2024).

<sup>18</sup> Basel, Staatsarchiv Basel-Stadt, STA H 43 22.

<sup>19</sup> Siehe <https://opendata.swiss/de/dataset/situationsplan-1862>. (Zugriff: 03.12.2024).



**Abb. 1** | Häuserbeschreibung in einer HGB-Karteikarte. Staatsarchiv Basel-Stadt, HGB 1 28/48, S. 41. Die mit Bleistift geschriebenen Notizen sind Querverweise auf die genannten Häuser. Staatsarchiv Basel-Stadt

eine als Falknerplan bekannte Katasterdarstellung von 1865.<sup>20</sup> Diese Pläne bieten eine verhältnismäßig präzise Darstellung des städtischen Raums und liegen georeferenziert vor. Somit existiert für eine Vielzahl der Adressen mit relativ wenig Aufwand eine belastbare Referenz, die sich visualisieren lässt.

Entsprach ein beschriebenes Haus nicht der Liegenschaft von 1862, vermerkten dies die HGB-Ersteller:innen, etwa als Teil einer Hausnummer oder als Kombination mehrerer Hausnummern. Neue Dossiers wurden angelegt, wenn sich Teilungen oder Zusammenlegungen von Häusern erkennen ließen. Wird etwa ein Haus mit „Blumenrain 10 u. Th. v. 8 neben 10“<sup>21</sup> beschrieben, so umfasste der damalige Besitz einen Teil der späteren Hausnummer 8 sowie die eigentliche Hausnummer 10. Die entsprechenden Angaben sind im Archivkatalog als zusätzliche Metadaten

<sup>20</sup> Siehe <https://opendata.swiss/en/dataset/grundbuchplan-basler-innerstadt-1865-72> (Zugriff: 03.12.2024).

<sup>21</sup> Basel, Staatsarchiv Basel-Stadt, HGB 1 28/56.

abgelegt, lassen sich jedoch nur bedingt regelbasiert auslesen. Das HGB enthält für den Zeitraum, der uns interessiert (also vor 1800) 3.984 Dossiers innerhalb des spätmittelalterlichen Mauerrings, davon 1.148, die einen Teil einer Hausnummer von 1862 beschreiben, und 603, die zusammengelegte Adressen betreffen.

Das Grundbuch- und Vermessungsamt des Kantons Basel-Stadt hat uns für 3.248 Dossiers Punktdaten überlassen, die auf einem Abgleich der Adressangaben basieren.<sup>22</sup> Weitere 324 Dossiers konnten wir auf die vom Vermessungsamt gelieferten Daten zurückführen, die restlichen 412 haben wir auf der Basis des Löffelplans manuell gesetzt. Die mehrere Hausnummern umfassenden Dossiers liegen dabei systematisch auf der tiefsten Hausnummer. Aufgrund der bewegten Geschichte der Häuser teilen sich gut 60 Prozent aller Häuserdossiers ihren Standort mit anderen Dossiers, die Punktdaten überlappen sich entsprechend vollständig. Die Überlappung der Punktdaten der Häuser zu reduzieren ist notwendig, da einerseits an einem Ort nur ein Haus stehen kann und andererseits die Darstellung sonst unleserlich wird.

### 3.2 Adresskorrektur

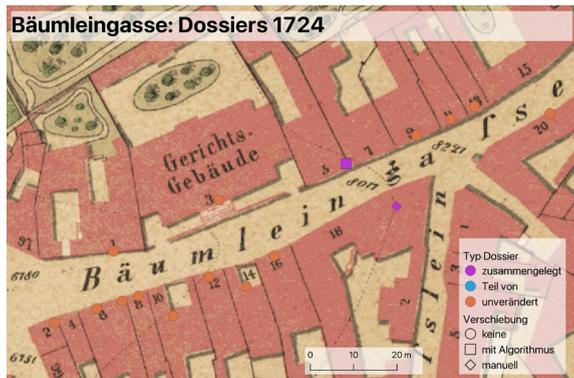
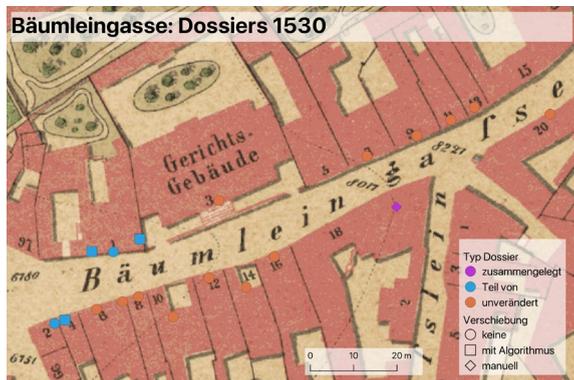
Dank der Informationen aus dem HGB können die Punkte von Häusern korrigiert und leserlicher georeferenziert werden. Bei geteilten Liegenschaften mit einer ‚neben‘-Angabe wird der Punkt des nebenan liegenden Hauses eruiert. Der Punkt wird dann ein Viertel der Strecke in Richtung Nachbarhaus gerückt.<sup>23</sup> Bei zusammengelegten Liegenschaften hingegen wird der Punkt in den geometrischen Mittelpunkt aller erwähnten Hausnummern gesetzt.

Das Resultat der Anwendung dieser Prinzipien in Zusammenhang mit der Information zu Anfang- und Enddaten von Häuserdossiers wird am Beispiel der Bäumleingasse dargestellt (Abb. 2). Ziel der neuen Lokalisierung ist eine erhöhte Lesbarkeit – die Überlappungen konnten größtenteils eliminiert werden. Ein Großteil dieser Arbeit ist regelbasiert automatisierbar, womit die händische Nacharbeit reduziert wird. In der Datenbank wird hinterlegt, aufgrund welcher Prinzipien ein Haus georeferenziert wurde und ob dies regelbasiert oder in händischer Nacharbeit geschehen ist. Somit können die Unsicherheiten (in Bezug auf die genaue Positionierung bzw. auf automatisch vorgenommene Verschiebungen) in Visualisierungen dargestellt werden. Das gewählte Vorgehen ist spezifisch an die Datenlage

---

<sup>22</sup> Oftmals erfolgte der Abgleich mit Hilfe der alten Hausnummer, die auf einer Durchnummerierung aller Häuser beruhte und von den HGB-Ersteller:innen auch notiert worden war.

<sup>23</sup> Der Wert ist willkürlich gewählt, es darf aber nicht die halbe Strecke sein, für den Fall, dass das Nebenhaus auch zeitweise geteilt ist (sonst würden diese Hälften wieder auf den gleichen Punkt zu liegen kommen).



**Abb. 2 |** Bäumleingasse. Oben: Darstellung der Adresspunkte anhand des Adressverzeichnisses von 1862, basierend auf den Daten des Grundbuch- und Vermessungsamts. Mitte und unten: Darstellungen der korrigierten Standpunkte von Dossiers, wie sie zu gewählten Zeitpunkten bestanden. Die Darstellung zeigt den Dossiertyp (Farbe) und die vorgenommene Verschiebung (keine, automatisiert oder händisch) an (siehe Legende). Löffelplan (<https://opendata.swiss/de/dataset/situationsplan-1862>). In QGIS erstellt und um Punktdaten ergänzt von Benjamin Hitz. CC BY.

gekoppelt, basiert aber auf allgemeinen Überlegungen und Prinzipien, die auch auf andere Katasterdaten übertragen werden können. Denn in aller Regel sind die aus Quellen erschließbaren historischen Daten nicht mit den verfügbaren Geodaten meist jüngeren Datums kompatibel. Auch bei lückenhafter Datenlage – etwa bei Nachbarschaftsreihen mit identifizierbaren und nicht lokalisierbaren Elementen – lohnt es sich, plausible Annahmen (als *educated guess*) zu treffen. Dabei ist die Reflexion über die Zielsetzung wichtig. Für viele räumliche Analysen ist die genaue Lokalisierung eines Hauses nicht zwingend erforderlich, während eine Darstellung der Daten ohne Hinweis auf die Unsicherheiten (einzelner) Daten zu vermeiden ist.

Wenn in der Folge also für die Basler Liegenschaften Punktdaten verwendet und dargestellt werden, sollte man sich stets der Unwägbarkeiten bewusst sein, die mit diesen Daten verbunden sind. Vor allem aber ist es wichtig, zu unterstreichen und zu dokumentieren, wie viele Datenquellen und verschiedene Verfahrensformen zusammenfließen, um diese Daten zu erstellen. Nur die Kombination von effizienter Kooperation und *Data Re-use* ermöglicht das Erstellen einer Datenbasis, auf deren Grundlage sich der Häuserbestand der Stadt Basel in einer langen Perspektive angemessen visualisieren und erforschen lässt.

## 4 Mit GIS visualisieren und quantifizieren

Die vorgestellten Punktdaten lassen sich nutzen, um Geovisualisierungen zu erstellen, die nicht nur die Existenz eines Gebäudes nachweisen, sondern sich auch mit einer Suche in den Exzerpten des HGB sowie deren Analyse koppeln lassen.<sup>24</sup> Die Darstellung wird – wie schon angesprochen – nicht als abschließende Visualisierung von Ergebnissen gesehen, sondern als zentraler Schritt in der Exploration des Datenbestandes und in der Nutzbarmachung der Daten für historische Fragestellungen aus der Sozial- und Wirtschaftsgeschichte. Deshalb folgt auch die Gliederung nicht inhaltlich den Analysen, die dargestellt werden, sondern der Exploration von Visualisierungs- und Analysemöglichkeiten.

### 4.1 Punkte in Raum und Zeit

Der erste Schritt auf dem Weg zur Geovisualisierung ist es, die vorliegenden Punktdaten als Punkte im Raum (und mit einem zeitlichen Filter) darzustellen. Als Beispiel soll die Verteilung von im HGB erwähnten Metzgern über die Stadt

---

<sup>24</sup> Hier möchten wir gerne den provisorischen Charakter der Analysen betonen, wobei es vorwiegend um die Diskussion der Darstellung von Daten und Suchresultaten geht.

gezeigt werden.<sup>25</sup> Das 15. Jahrhundert wurde als Zeitraum gewählt, da zuvor die Anzahl der Suchtreffer deutlich kleiner war, danach hingegen die Streuung weniger auffällig. Die reine Erwähnung eines Metzgers belegt diesen noch nicht als Besitzer oder Bewohner der betreffenden Liegenschaft, er konnte auch Zinsgläubiger, Kläger oder Nachbar sein (Abb. 3).<sup>26</sup>

Schon diese erste, simple Darstellung lässt Schlüsse zu. So lässt sich etwa eine Konzentration von Metzger-Erwähnungen im Südwesten des Kornmarkts (des heutigen Marktplatzes) erkennen. Insgesamt ist aber eine Darstellung in dieser Form noch von beschränkter Aussagekraft, da die Rolle der Metzger nicht explizit geklärt ist. Die vielen Überlappungen von Punkten ergeben keinen gesicherten Eindruck, sondern sind aufgrund der Überschneidungen sehr willkürlich. Erst durch die darauf folgenden konsequenten Auswertungen der Häufungen lassen sich Auswertungen stabilisieren, da sie eine (auch automatische) Präzisierung ermöglichen.

Um die Übersicht über die Daten zu verbessern, basieren die nun folgenden Analysen und Visualisierungen auf verschiedenen Formen der Aggregation von Daten. Diese hat auch den Vorteil, dass die genaue Verortung im Regelfall an Bedeutung verliert beziehungsweise allfällige Fehler weniger ins Gewicht fallen.

## 4.2 Heatmaps: Häufigkeiten identifizierbar machen

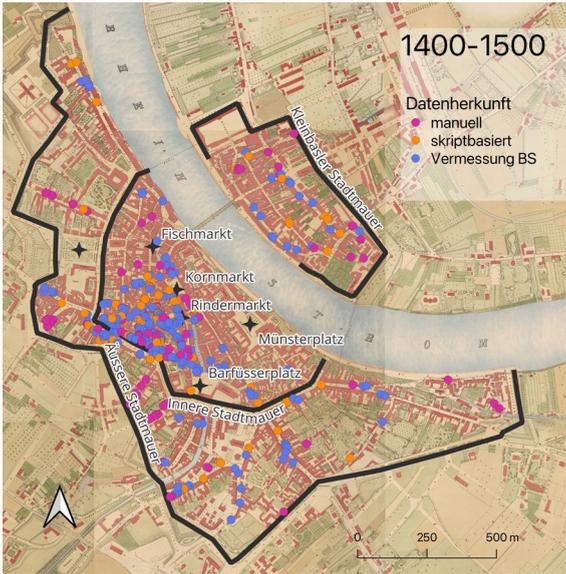
Die Streuung von Punkten lässt sich mit sogenannten Heatmaps einfacher interpretieren. Dabei werden Punkte nicht mehr einzeln dargestellt, sondern aggregiert, was die Unsicherheit bezüglich des genauen Standortes reduziert. In einer Heatmap sind Gebiete mit hoher Konzentration dunkler eingefärbt. Um die Vergleichbarkeit mit anderen Zeitpunkten und verschiedenen Skalen zu ermöglichen, wurden für die Heatmaps in diesem Artikel jeweils identische Einstellungen gewählt (Abb. 4).<sup>27</sup>

Die umgesetzte Heatmap der Metzger kontrastiert stark mit einer Punktwolke, denn sie zeigt die schon oben erkannte Massierung. Weil die übrigen Erwähnungen von Metzgern eine deutlich kleinere räumliche Dichte aufweisen, verschwinden sie praktisch. Die klare Konzentration von Metzgern im Südwesten des Kornmarkts

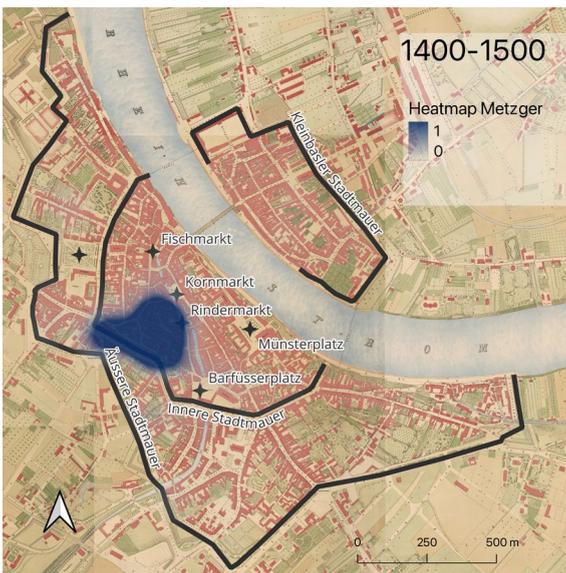
<sup>25</sup> Datengrundlage: Suche nach „transkript ~, ^.\*[Mm]e[t]?zger.\*\$“, Suchresultate gefiltert nach Zeitraum 1400–1500, n = 860.

<sup>26</sup> Auf die Erkennung von Ereignissen und Rollenzuweisungen wird in diesem Beitrag nicht eingegangen.

<sup>27</sup> Hier konkret: Farbrampe Dunkelblau zu Transparent, 85% *opacity*, 120 m Radius in *map units*, 80 *maximum value*. Diese Einstellung sorgt für eine Vergleichbarkeit mit anderen Zeitpunkten, indem der *maximum value* festgelegt (und nicht für jede Heatmap neu berechnet) wird. Die Definition einer Größe in Karteneinheiten sorgt dafür, dass die Darstellung der Heatmap nicht abhängig ist vom Kartenmaßstab.



**Abb. 3 |** Verteilung aller Erwähnungen von Metzgern im Zeitraum 1400–1500 (n = 860). Löffelplan (<https://opendata.swiss/de/dataset/situationsplan-1862>). In QGIS erstellt und um Punktdaten ergänzt von Benjamin Hitz. CC BY.



**Abb. 4 |** Heatmap aller Erwähnungen von Metzgern im Zeitraum 1400–1500. Löffelplan (<https://opendata.swiss/de/dataset/situationsplan-1862>). In QGIS erstellt und um Heatmap ergänzt von Benjamin Hitz. CC BY.

hing nur bedingt mit ihrem Beruf zusammen, denn der Standort der Fleischbänke, wo der Verkauf stattfinden musste, lag außerhalb der dichten Zone unmittelbar nördlich des Kornmarkts. Die Heatmap bestätigt die bekannte Sozialtopographie der Metzgerzunft, wie sie anhand von Steuerlisten für das 15. Jahrhundert schon beschrieben wurde.<sup>28</sup> Im Vergleich zu den nur punktuell vorhandenen Steuerdaten erlaubt unsere Datenquelle jedoch eine dynamische Analyse, etwa um Veränderungen zu erkennen. So kam im frühen 16. Jahrhundert ein erhöhtes Vorkommen an Metzgern in der Spalenvorstadt dazu, das bislang unbekannt war.

Ebenfalls geeignet zur Darstellung in Form von Heatmaps erscheinen uns *frönungen* (neuhochdeutsch als ‚Pfändung‘ zu übersetzen), ein juristisches Verfahren, bei dem Gläubiger:innen – darunter oft kirchliche Institutionen – eine Schuld einforderten, die an eine Liegenschaft gebunden war. Wurde eine *frönung* dreimal vor Gericht bestätigt (jeweils mit zwei Wochen Wartefrist), konnte die Liegenschaft gegen den Willen der Besitzer:innen erworben werden. Nur wenige angestrebte Verfahren wurden bis zum Ende durchgeführt. Meist bemühten sich die Schuldner:innen, die Beschlagnahmung ihrer Liegenschaft abzuwenden.<sup>29</sup> In Abbildung 6 sind die vielen angedrohten Beschlagnahmungen dargestellt. Die eindeutige Bezeichnung des Verfahrens ermöglicht es, Karteikarten zu identifizieren, die eine *frönung* (in der Regel als Partizip von *frönen*) erwähnen: Gesucht wurde im Haupttext der Karteikarte nach Flexionsformen und Varianten des Verbs *frönen*.<sup>30</sup> Die Gesamtzahl von Treffern belief sich im Zeitraum von 1353 bis 1805 auf 9.770.

*frönungen* lassen sich überlieferungsbedingt erst seit etwa 1400 erkennen. Sie verteilten sich im Zeitverlauf in ähnlichen Mustern. 1500 bis 1550 wurden vergleichsweise weniger Liegenschaften beschlagnahmt als in den Jahren davor, aber doch in vergleichbarer Streuung. Ein Grund dafür könnte die Reformation gewesen sein, denn die Überführung der Klöster und ihrer Zinsansprüche in die städtische Verwaltung könnte zu einer Abnahme der Verfahren geführt haben (Abb. 5).

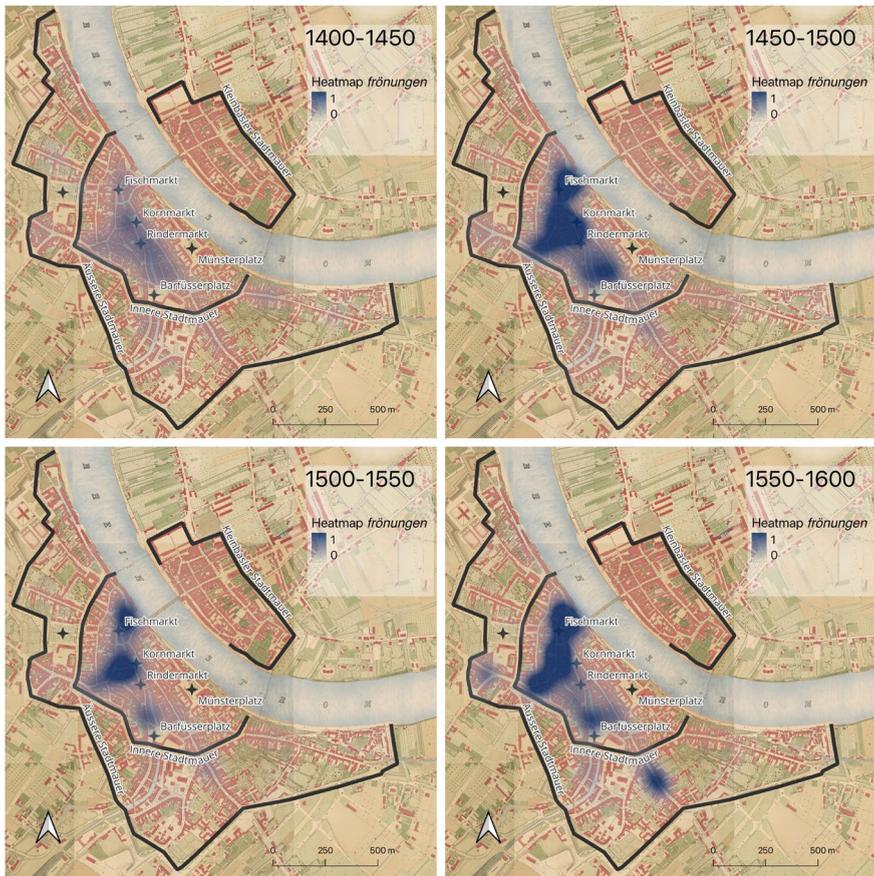
In den absoluten Zahlen der Heatmap überwiegt das dichtbebaute Zentrum. Das ist erstaunlich, da Schuldeintreibungsverfahren eher mit Armut assoziiert werden. Obwohl die Basler Sozialtopographie für das frühe 16. Jahrhundert anhand des HGB untersucht wurde, liegen keine Resultate vor, die sich in ihrer Granularität mit denjenigen der von uns erstellten Heatmaps vergleichen ließen.<sup>31</sup> Es scheint jedoch, dass sich die dichten Gebiete der Heatmaps mit den eher gehobenen Quartieren, in denen viele reichen Bürger:innen wohnten, zumindest

28 Vgl. Simon-Muscheid 1988, S. 98–103. Zu städtischen Sozialtopographien allgemein siehe Arnaud 2018; Gerber 2001; Rüthing 1986; Schoch 1997.

29 Vgl. Hitz 2022, S. 186f.

30 SQL-Komponente der Suche: `layout_typ like ‚paragraph‘ and transkript ~ ‚^.*gefr[oö][h]?n[n]?t.*$‘`.

31 Vgl. Füglistner 1981, Kap. II.



**Abb. 5 |** Heatmaps der gefundenen *frönungen* in Abschnitten von 50 Jahren. 1400–1600 (1400–1450: n = 1075, 1450–1500: n = 1721, 1500–1550: n = 1182, 1550–1600: n = 1472). Löffelplan (<https://opendata.swiss/de/dataset/situationsplan-1862>). In QGIS erstellt und um Heatmap ergänzt von Benjamin Hitz. CC BY.

überschneiden. Es gibt Hinweise darauf, dass auch Reiche *frönungen* unterworfen waren,<sup>32</sup> wobei es weiter zu prüfen gilt, inwiefern die gewählte Auswertungsstrategie dazu geeignet ist.

<sup>32</sup> Vgl. Hitz 2022, S. 192–196.

### 4.3 Den Raum einteilen: Hexagone als Analysegröße

Für verschiedene Analyse Zwecke kann es sich lohnen, den untersuchten Raum mit einem Raster zu überziehen, dessen Einheiten dazu dienen, Daten zu aggregieren. Für die Analyse der *frönungen* wurde eine Gliederung in Hexagone mit einer Breite von 150 Metern gewählt – alternativ könnte auch mit anderen Formen, etwa Quadraten, gearbeitet werden. Die Hexagone haben den Vorteil, dass sie nicht gegebenen Vorstellungen räumlicher Gliederung folgen, sondern die Visualisierung quasi verfremden und einen neutralen Blick auf räumliche Strukturen ermöglichen. Die Punktdaten werden für diese Analysen einem Hexagon zugewiesen, das entsprechend eingefärbt wird, etwa indem die Färbung für eine höhere Anzahl von Punktdaten dunkler ausfällt.

Der wichtigste Vorteil dieser Methode besteht darin, dass wir jegliche Suchtreffer in Bezug setzen können zu anderen Daten innerhalb des Hexagons. Für die folgenden Beispiele ist das die Anzahl an Dossiers, die zum jeweiligen Zeitpunkt aktiv waren. So lässt sich die Dichte nicht mehr nur räumlich, sondern relativ zu den bestehenden Liegenschaften berechnen, was präzisere Aussagen zur räumlichen Streuung ermöglicht. Die Dichte berechnet sich dabei aus der durchschnittlichen Anzahl an Treffern pro Dossier.

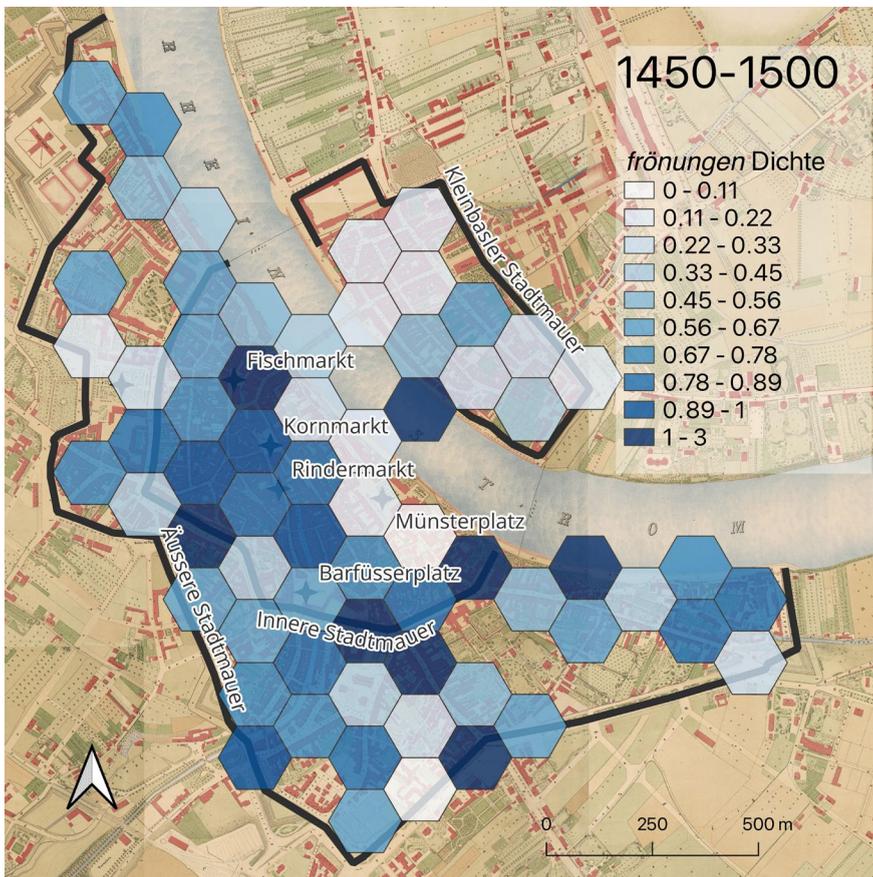
Die Darstellung der *frönungen* mit der beschriebenen Methode ist ein interessanter Anwendungsfall. Im Vergleich zur Heatmap relativiert sich in dieser Darstellung die Konzentration der *frönungen* im Stadtzentrum um den Kornmarkt. Wohl ist auch dort der Anteil hoch, aber es gibt gewisse Gebiete in den Vorstädten mit ähnlich hohen Werten, während zentrumsnahe Regionen wie der Münsterhügel am Rheinufer mit sehr tiefen Werten auffallen (Abb. 6).

### 4.4 Cluster: den Raum neu gliedern

Abschließend sei noch ein alternatives Vorgehen mit entgegengesetztem Ansatz vorgestellt. Statt nämlich die Daten in vorgegebenen Gliederungen zu analysieren, lassen sich aus den Daten selbst neue Gliederungen erschließen. Zur Anwendung kommt ein Algorithmus, der Cluster naher Nachbarn berechnet (*Nearest Neighbours Clusters Map*<sup>33</sup>). Dieser wurde ursprünglich für die Analyse von Kriminalstatistiken entwickelt; er identifiziert Hotspots von Ereignissen, indem er zunächst für alle Ereignisse – hier unsere Suchtreffer – nahe Nachbarn anhand einer definierten Suchdistanz auswählt. Das Ereignis mit den meisten Nachbarn bildet den ersten Hotspot, und dieses wird aus der Verteilung entfernt. Der Prozess

---

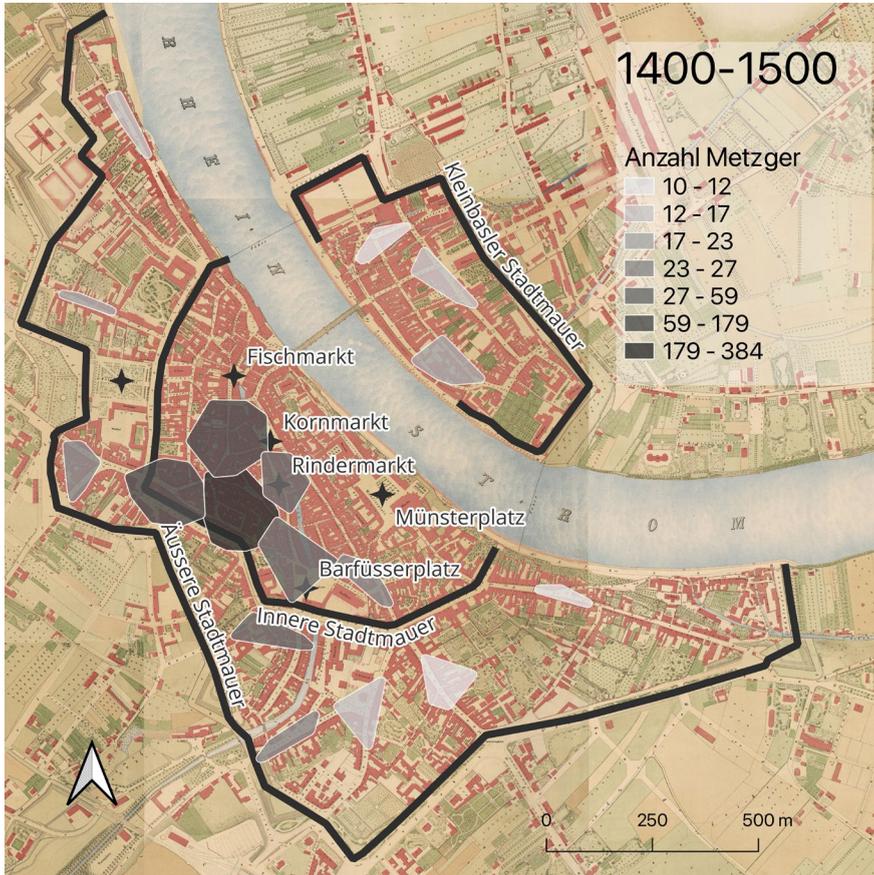
<sup>33</sup> Siehe <https://analysecriminelle.org/main/#nearest-neighbours-clusters-map> (Zugriff: 03.12.2024).



**Abb. 6** | frönungen 1450–1500, relativ zur Anzahl aktiver Dossiers pro Hexagon. Löffelplan (<https://opendata.swiss/de/dataset/situationsplan-1862>). In QGIS erstellt und um Hexagon-Layer ergänzt von Benjamin Hitz. CC BY.

wird wiederholt, bis keine Ereignisse mehr übrig sind, die die Mindestanzahl an Nachbarn erreichen. Danach wird der Schwerpunkt jedes Clusters berechnet. Um Überlappungen zu vermeiden, werden die Entfernungen zu diesen Schwerpunkten neu bestimmt. Schließlich wird die äußere Begrenzung eines jeden Clusters verwendet, um die Hotspots als Polygone darzustellen.

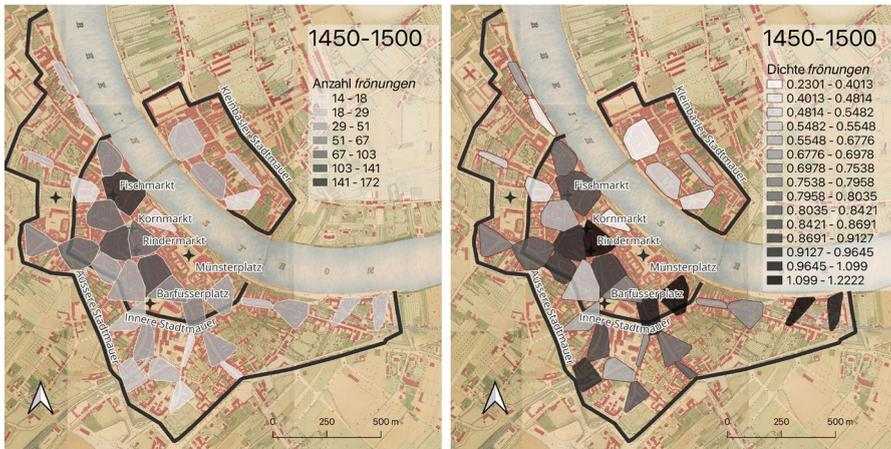
Die Clusterkarte der Metzger zeigt als Zentrum der Erwähnung von Metzgern denselben Bereich wie der dichte Teil der Heatmap. Deutlicher als in der Heatmap wird aber die Strukturierung und Ansammlung um dieses Zentrum herum. Zudem ist zu erkennen, dass es angrenzende Cluster mit fast ähnlich hohen Zahlen von Metzgern gibt, vor allem aber findet diese Methode noch Auffälligkeiten



**Abb. 7 |** Nearest Neighbours Clusters Map der Metzger, 1400-1500. Löffelplan (<https://open-data.swiss/de/dataset/situationsplan-1862>). In QGIS erstellt und um Clustering-Layer ergänzt von Benjamin Hitz. CC BY.

in Gebieten, die von der Heatmap, die auf besonders dichte Gebiete fokussiert, weitgehend weiß gelassen wurden. Einzelne Erwähnungen mit nur sehr wenigen Nachbarn in der geforderten Distanz fallen hingegen weg (Abb. 7).

Das gleiche Clustering haben wir für den Zeitraum 1450 bis 1500 auf die Vorkommen von *frönungen* angewandt. Im linken Teil der Abbildung ist das Suchresultat nach Anzahl der Punkte im Cluster eingefärbt (wie oben die Metzger). Es zeigt sich, dass fast das ganze bebaute Gebiet Teil eines Clusters war, mit Ausnahme gewisser Teile Kleinbasels, was wohl auch auf eine Überlieferungslücke zurückzuführen ist, und dem schon erwähnten Münsterhügel. Mit den Clustern können wir analog zum Vorgehen bei den Hexagonen die Dichte relativ zu den



**Abb. 8 |** Nearest Neighbour Cluster Maps der *frönungen*, 1450–1500. Links: Färbung der Cluster nach Anzahl der *frönungen* (je dunkler, desto höher), rechts: Färbung der Cluster nach Verhältnis der *frönungen* zu den vorhandenen Dossiers im Cluster (je höher die Dichte, desto dunkler). Löffelplan (<https://opendata.swiss/de/dataset/situationsplan-1862>), in QGIS erstellt und um Clustering-Layer ergänzt von Benjamin Hitz. CC BY.

bestehenden Dossiers im entsprechenden Polygon darstellen (rechts). Dadurch relativiert sich die Bedeutung des großen und dichten Clusters am Brückenkopf des Rheins, und andere, kleinere und zum Teil peripherere Cluster treten in den Vordergrund. *frönungen* treten uns in dieser letzten und wohl am besten lesbaren Darstellung als ein äußerst komplexes Phänomen entgegen: Sie sind sowohl in wirtschaftlich dynamischen Zentrumsregionen, wo viele Wohlhabende wohnten, als auch in gewissen Teilen der städtischen Peripherie häufig anzutreffen. Die Gebiete mit geringer Dichte an *frönungen* sind oft gegensätzlich ausgeprägt, wie beispielsweise ein Vergleich zeigt zwischen dem Münsterhügel, auf dem die Geistlichkeit sehr präsent war, und der St. Johannsvorstadt nördlich der Rheinbrücke, in der viele Fischer und Schiffer lebten. Die Darstellung wirft somit mindestens so viele Fragen auf, wie sie Antworten gibt (Abb. 8).

## 5 Schluss

Die vorliegende Untersuchung zeigt, wie durch die Kombination digitaler Methoden und unterschiedlicher Datenquellen neue Explorationsmöglichkeiten für die städtische Wirtschaftsgeschichte der Vormoderne gewonnen werden können. Unsere Arbeit demonstriert, dass die Nutzung von GIS und die Entwicklung von Datenvisualisierungen nicht nur die Analyse historischer Daten erheblich

erleichtern, sondern auch neue Perspektiven auf wirtschaftliche und soziale Strukturen eröffnen können und somit neue Formen der Thesenbildung unterstützen.

Unsere Methodik zur geographischen Visualisierung und Analyse, einschließlich der Nutzung von Heatmaps, Hexagon-Rastern und Cluster-Algorithmen, verdeutlicht nicht nur die räumliche Verteilung von Berufsnennungen und rechtlichen Verfahren, sondern hilft auch, die Dynamiken der städtischen Entwicklung zu erfassen. Die räumliche Darstellung der Nennung von Metzgern und *frönungen* zeigt, dass solche Visualisierungen nicht nur bestehende Annahmen bestätigen, sondern auch neue Fragestellungen aufwerfen und bislang unbekannte Muster aufdecken können. Aus diesem Grund ist es zentral, die visuelle Darstellung als Form einer intermediären Analyse zu verstehen, die zu Entdeckungen (im Sinne der *Serendipity*)<sup>34</sup> beiträgt und Vorstellungen und Annahmen datengestützt hinterfragt und herausfordert. Dabei zeigt sich auch, dass die Form der Visualisierung mögliche Interpretationen beeinflusst, weshalb selten eine Visualisierung ausreicht, sondern mehrere parallel verglichen werden sollten.

Ein weiterer wichtiger Aspekt war die Auseinandersetzung mit Unsicherheiten und Fehlerquellen in den historischen Daten. Die sorgfältige Dokumentation der Datenentstehung und die Markierung der damit verbundenen Unsicherheiten sind entscheidend für die Interpretierbarkeit der Visualisierungen. Die scheinbar objektivierte Darstellung in Visualisierungen sollte durchbrochen und heuristische und quellenkritische Herausforderungen sollten markiert werden. Nicht zuletzt, weil durch dynamische Werkzeuge und interaktive Umsetzungen mit wenigen Zeilen Programmiercode umfangreiche Datensätze visualisiert werden können. Auch wenn die Aufbereitung und Visualisierung von Daten wie den hier präsentierten erheblichen Aufwand zeitigen und einen zusätzlichen Schritt in der Auswertung nach sich ziehen, legitimieren potentielle Erkenntnisgewinne notwendige Mittel und investierte Arbeitszeit durchaus.

## Literaturverzeichnis

### Quellen

- Adressbuch der Stadt Basel von 1862. Basel, Staatsarchiv Basel-Stadt, STA H 43 22. (im digitalen Lesesaal): <https://dls.staatsarchiv.bs.ch/records/1016781>.
- HGB = Historisches Grundbuch Basel. Basel, Staatsarchiv Basel-Stadt. Online

---

34 Zur Diskussion der *Serendipity* in den *Digital Humanities* siehe Verhoeven 2016.

## Forschungsliteratur

- Arnaud, Colin:** Topographien des Alltags. Bologna und Strassburg um 1400 (Europa im Mittelalter 28). Berlin 2018. <https://doi.org/10.1515/9783110545906> (Zugriff: 03.12.2024).
- Bavel, Bas J. P. van:** Markets for Land, Labor, and Capital in Northern Italy and the Low Countries, Twelfth to Seventeenth Centuries. In: *The Journal of Interdisciplinary History* 41 (2011), S. 503–531. [https://doi.org/10.1162/JINH\\_a\\_00201](https://doi.org/10.1162/JINH_a_00201) (Zugriff: 03.12.2024).
- Birnbaum, David J., Sheila Bonde u. Mike Kestemont:** The Digital Middle Ages: An Introduction. In: *Speculum* 92 (2017), S. 1–38. <https://doi.org/10.1086/694236> (Zugriff: 03.12.2024).
- Drucker, Johanna:** Humanities Approaches to Graphical Display. In: *Digital Humanities Quarterly* 5 (2011). <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html> (Zugriff: 03.12.2024).
- Drucker, Johanna:** Graphical Approaches to the Digital Humanities. In: Susan Schreibman, Ray Siemens u. John Unsworth (Hgg.): *A New Companion to Digital Humanities*. Chichester 2015, S. 238–250. <https://doi.org/10.1002/9781118680605.ch17> (Zugriff: 03.12.2024).
- Ertl, Thomas:** Wien 1448. Steuerwesen und Wohnverhältnisse in einer spätmittelalterlichen Stadt. Wien 2020.
- Fouquet, Gerhard u. Kurt Andermann (Hgg.):** Zins und Gült. Strukturen des ländlichen Kreditwesens in Spätmittelalter und Frühneuzeit. Ependorf 2016.
- Füglister, Hans:** Handwerksregiment: Untersuchungen und Materialien zur sozialen und politischen Struktur der Stadt Basel in der ersten Hälfte des 16. *Jahrhunderts* (Basler Beiträge zur Geschichtswissenschaft 143). Basel 1981.
- Gerber, Roland:** Gott ist Burger zu Bern. Eine spätmittelalterliche Stadtgesellschaft zwischen Herrschaftsbildung und sozialem Ausgleich. Weimar 2001.
- Gerber, Roland:** Umgestaltung und Neubeginn: der wirtschaftliche und soziale Wandel Berns im 15. Jahrhundert. In: *Zeitschrift für historische Forschung* 29 (2002), S. 161–188.
- Gilomen, Hans-Jörg:** Wirtschaftsgeschichte des Mittelalters. München 2014.
- Hammel, Rolf:** Der Immobilienmarkt im Spätmittelalter und in der frühen Neuzeit am Beispiel der Hansestadt Lübeck. In: Uwe Greve (Hg.): *Von der Felshöhle zum Wolkenkratzer. Zur Entwicklung des Hauswesens und Grundbesitzes im Abendland von den Anfängen bis zur Gegenwart*. Husum 1990, S. 107–132.
- Harkes, Rachael C.:** Sociological Approaches and the Urban History of Medieval England. *Research Trends and New Perspectives* (2017–2022). In: *Urban History* 49 (2022), S. 648–656. <https://doi.org/10.1017/S0963926822000293> (Zugriff: 03.12.2024).
- Heidmann, Frank:** Interaktive Karten und Geovisualisierungen. In: Wibke Weber, Michael Burmester u. Ralph Tille (Hgg.): *Interaktive Infografiken*. Berlin, Heidelberg 2013, S. 39–69. [https://doi.org/10.1007/978-3-642-15453-9\\_3](https://doi.org/10.1007/978-3-642-15453-9_3) (Zugriff: 03.12.2024).
- Hitz, Benjamin:** Ein Netz von Schulden. Schuldbeziehungen und Gerichtsnutzung im spätmittelalterlichen Basel (Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte. Beiheft 256). Stuttgart 2022.

- Howell, Martha C.:** Commerce before Capitalism in Europe, 1300–1600. Cambridge 2010.
- Jessop, Martyn:** Digital Visualization as a Scholarly Activity. In: Literary and Linguistic Computing 23 (2008), S. 281–293. <https://doi.org/10.1093/llc/fqn016> (Zugriff: 03.12.2024).
- Krauer, Rezia:** Die Beteiligung städtischer Akteure am ländlichen Bodenmarkt. Die Region St. Gallen im 13. und 14. Jahrhundert. Zürich 2018. <https://doi.org/10.5167/uzh-164765> (Zugriff: 03.12.2024).
- Lorenzen-Schmidt, Klaus-Joachim:** Die Immobilienmärkte von Lübeck und Hamburg 1455–1599. In: Michael Hundt u. Jan Lokers (Hgg.): Hanse und Stadt. Akteure, Strukturen und Entwicklungen im regionalen und europäischen Raum. Lübeck 2014, S. 313–321.
- Muehlberger, Guenter, Louise Seaward, Melissa Terras u. a.:** Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study. In: Journal of Documentation 75 (2019), S. 954–976. <https://doi.org/10.1108/JD-07-2018-0114> (Zugriff: 03.12.2024).
- Owens, J. B.:** Toward a Geographically-Integrated, Connected World History: Employing Geographic Information Systems (GIS). In: History Compass 5/6 (2007), S. 2014–2040. <https://doi.org/10.1111/j.1478-0542.2007.00476.x> (Zugriff: 03.12.2024).
- Perrone, Sean T., Adam Franklin-Lyons, David Gary Shaw u. a.:** Visualization, Mapping, and the History of Mobility in the Middle Ages. The Historian 84 (2022), S. 504–538. <https://www.tandfonline.com/doi/abs/10.1080/00182370.2023.2262868> (Zugriff: 03.12.2024).
- Perrone, Sean T. u. Carol Traynor:** Mapping the Way of St. James: GIS Technology, Spatial History, and the Middle Ages. In: Church History and Religious Culture 101 (2021), S. 3–32. <https://doi.org/10.1163/18712428-bja10013> (Zugriff: 03.12.2024).
- QGIS Association: QGIS.org. <http://www.qgis.org> (Zugriff: 03.12.2024).
- Rüthing, Heinrich:** Höxter um 1500. Analyse einer Stadtgesellschaft. Paderborn 1986.
- Schoch, Willi:** Die Bevölkerung der Stadt St. Gallen im Jahre 1411. Eine sozialgeschichtliche und sozialtopographische Untersuchung. St. Gallen 1997.
- Schwinges, Rainer Christoph:** Das Repertorium Academicum Germanicum (RAG). Ein digitales Forschungsvorhaben zur Geschichte der Gelehrten des Alten Reiches (1250–1550). In: Jahrbuch für Universitätsgeschichte 16 (2015), S. 215–232.
- Sidiropoulos, George:** The Contribution of the Cartographic Visualisation in Understanding the Historical Geography of an Old Land Property Regime in Mouth Athos (Karyes). In: The Cartographic Journal 47 (2010), S. 270–277. <https://doi.org/10.1179/000870410X12786821061576> (Zugriff: 03.12.2024).
- Simon-Muscheid, Katharina:** Basler Handwerkszünfte im Spätmittelalter. Zunftinterne Strukturen und innerstädtische Konflikte. Bern 1988.
- Tufte, Edward R.:** The Visual Display of Quantitative Information. Cheshire 2001.
- Verhoeven, Deb:** As Luck Would Have It. Serendipity and Solace in Digital Research Infrastructure. In: Feminist Media Histories 2 (2016), S. 7–28. <https://doi.org/10.1525/fmh.2016.2.1.7> (Zugriff: 03.12.2024).

**Vetch, Paul, Catherine Clarke u. Keith**

**Lilley:** Between Text and Image: Digital Renderings of a Late Medieval City. In: Brent Nelson u. Melissa Terras (Hgg.):

Digitizing Medieval and Early Modern Material Culture. New Technologies in Medieval and Renaissance Studies. New Mexico 2012.

# Im Netz mittelhochdeutscher Wörter Das ‚Mittelhochdeutsche Wörterbuch‘ und seine Quellen

## Kontakt

**Patrick D. Brookshire**,  
Mittelhochdeutsches Wörterbuch/  
Digitale Akademie,  
Akademie der Wissenschaften und der  
Literatur Mainz,  
Geschwister-Scholl-Str. 2,  
55131 Mainz,  
patrick.brookshire@adwmainz.de  
 <https://orcid.org/0000-0002-5843-7577>

**Dr. Jonas Richter**,  
Mittelhochdeutsches Wörterbuch,  
Niedersächsische Akademie der  
Wissenschaften zu Göttingen,  
Geiststraße 10,  
37073 Göttingen,  
jonas.richter@adwgoe.de  
 <https://orcid.org/0000-0002-2924-6026>

**Abstract** This paper illustrates how existing interdependencies between various Middle High German dictionaries have, over time, lead to a dense network with ‘Mittelhochdeutsches Wörterbuch’ (‘Middle High German Dictionary’) at its core. This is due to the fact that it offers multiple interfaces that can be used to connect Middle High German words between different projects. One such interface is the lemma list, which has been around since the early days of the project, and another more recent one is the Federated Content Search of Text+ (a consortium of the National Research Data Infrastructure in Germany). In addition, we show how the digitally available primary source corpus can be reused by presenting a tool for cooccurrence analyses which also considers metadata, as well as a pilot study on Named Entity Recognition. With all of this, we not only highlight existing cooperation and collaboration efforts but also want to foster future interdisciplinary research to let this network grow even more.

**Keywords** Corpus Linguistics; Digital Medieval Studies; Federated Content Search; Middle High German; Named Entity Recognition

## 1 Hintergrund

Im Sommer 1867, nur ein Jahr nach der Vollendung des ‚Mittelhochdeutschen Wörterbuchs‘ von BENECKE-MÜLLER-ZARNCKE (BMZ), bat der Verleger Salomon HIRZEL Matthias LEXER, ein Handwörterbuch auszuarbeiten, das zugleich als Supplement und als alphabetischer Index zum BMZ dienen sollte.<sup>1</sup> Für die Ausarbeitung benötigte LEXER ein Vielfaches der Zeit, die HIRZEL 1869 veranschlagt hatte, und das ‚Handwörterbuch‘ überstieg auch den geplanten Umfang deutlich. Hinzu kamen 400 Spalten mit Nachträgen, die LEXER seinem ‚Handwörterbuch‘ als Anhang beifügte.<sup>2</sup> In der Altgermanistik etablierten sich BMZ und LEXER als Zweigespann, mit denen das Netz mittelhochdeutscher Wörter zu erkunden war – wenn auch nicht ohne Klagen darüber, dass der Zugriff über die zwei Alphabete im LEXER (Hauptteil und Nachträge) sowie im Nebeneinander der zwei Wörterbücher umständlich sei.<sup>3</sup> Um die Verstrickungsgefahr in diesem Netz zu verringern, wurden 1997 bis 2000 DFG-finanziert die beiden Wörterbücher sowie das ‚Findebuch‘<sup>4</sup> retrodigitalisiert, miteinander verlinkt und als ‚Mittelhochdeutsche Wörterbücher im Verbund‘ auf CD-ROM und im Internet veröffentlicht. Dieser Verbund ist mittlerweile im ‚Wörterbuchnetz‘ aufgegangen.<sup>5</sup>

Trotz dieser jüngeren, elektronischen Entwicklung sollte aber nicht vergessen werden, dass die systematische, werkübergreifende Vernetzung des mittelhochdeutschen Wortschatzes mit Matthias LEXER begann, der in seinem Handwörterbuch konsequent nachwies, wo im BMZ der entsprechende Eintrag zu finden ist. In gewisser Weise kann man das von HIRZEL initiierte und von LEXER ausgearbeitete ‚Handwörterbuch‘ auch als einen kleinen Schritt in Richtung der FAIR-Prinzipien<sup>6</sup> betrachten, da er die Auffindbarkeit (*Findability*) der BMZ-Daten erhöhte und sie für seine eigenen Wörterbucheinträge nachnutzte bzw. darauf aufbaute (*Reusability*).

Die digitalen Möglichkeiten erlauben aber eine noch engere Verzahnung, als sie zu dieser Zeit erreicht werden konnte. In der erwähnten Retrodigitalisierung Ende der 1990er Jahre wurden die Einträge von BMZ, LEXER und ‚Findebuch‘ mit Hyperlinks verbunden, außerdem die Quellensiglen mit NELLMANNS ‚Verzeichnis

---

1 Lexer 1872, S. v.

2 Vgl. Gärtner 1993.

3 Vgl. Nellmann 1991, S. 255.

4 Vgl. Gärtner, Gerhardt, Jährling u. a. 1992.

5 <https://woerterbuchnetz.de/> (Zugriff: 11.03.2025).

6 ‚FAIR‘ (Wilkinson, Dumontier, Aalbersberg u. a. 2016) steht für ‚Findability‘, ‚Accessibility‘, ‚Interoperability‘ und ‚Reusability‘.

der benutzten Werke<sup>7</sup> verknüpft.<sup>8</sup> Im neuen ‚Mittelhochdeutschen Wörterbuch‘ (MWB), das nach Vorarbeiten in den 1990ern seit 2000 als Langzeitprojekt im Akademienprogramm geführt wird und 2032 abgeschlossen sein soll, kommen weitere Verlinkungen hinzu.<sup>9</sup>

Einen besonderen Mehrwert stellt nach GÄRTNER und PLATE die Verknüpfung von Belegstellen in den Wörterbuchartikeln mit Volltexten dar:<sup>10</sup> Viele Quellen des MWB sind elektronisch ins System eingepflegt, so dass die verwendeten Belegzitate mit dem Volltext verknüpft bleiben. Bei der Nutzung des Online-Wörterbuchs kann also bei Bedarf eine Referenz angeklickt werden, um den Kontext des Belegs anzusehen. Die Verlinkung ist dabei bidirektional: Neben dem Volltext wird angezeigt, in welchen Artikeln die jeweiligen Textstellen zitiert sind.

Als *born-digital* Wörterbuch mit hybrider Publikationsform (elektronisch und gedruckt) bemüht sich das Projekt, neben der Hauptarbeit, dem Erstellen der Wörterbuchartikel, seine Inhalte auch als nachnutzbare und vernetzungsfähige Daten aufzubereiten. Dazu dienen die Verwendung etablierter Standards der Datenmodellierung und die Bereitstellung verschiedener Schnittstellen (API), die einen automatisierten Zugriff auf die Daten des Projektes (und nicht nur auf die Wörterbuchartikel) ermöglichen. Durch digitale Methoden können interoperable Daten deutlich einfacher über Projektgrenzen hinaus geteilt oder über föderierte Architekturen gemeinsam durchsucht werden, ohne dass die Datensouveränität und die Sichtbarkeit der Einzelprojekte verlorengehen.

Im folgenden Abschnitt wird zunächst auf Aspekte der Datenmodellierung innerhalb des MWBs eingegangen. Anschließend werden im 3. Abschnitt bereits existierende Schnittstellen zu anderen Projekten am Beispiel der Gesamtlemma-Liste des Projekts sowie der föderierten Suche von Text+ vorgestellt. Als weitere Nachnutzungsmöglichkeiten von MWB-Daten thematisiert der 4. Abschnitt einerseits ein Tool zur Kookkurenzsuche und andererseits eine Pilotstudie zur automatischen Eigennamenerkennung in mittelhochdeutschen Texten.

## 2 Datenmodellierung

Die beiden zentralen Ressourcen des MWB (Wörterbuchartikel und Primärquellen im Volltext) sind nach den Richtlinien der *Text Encoding Initiative* (TEI) in XML codiert und damit in hohem Maß nachnutzbar und interoperabel. Die Wortartikel

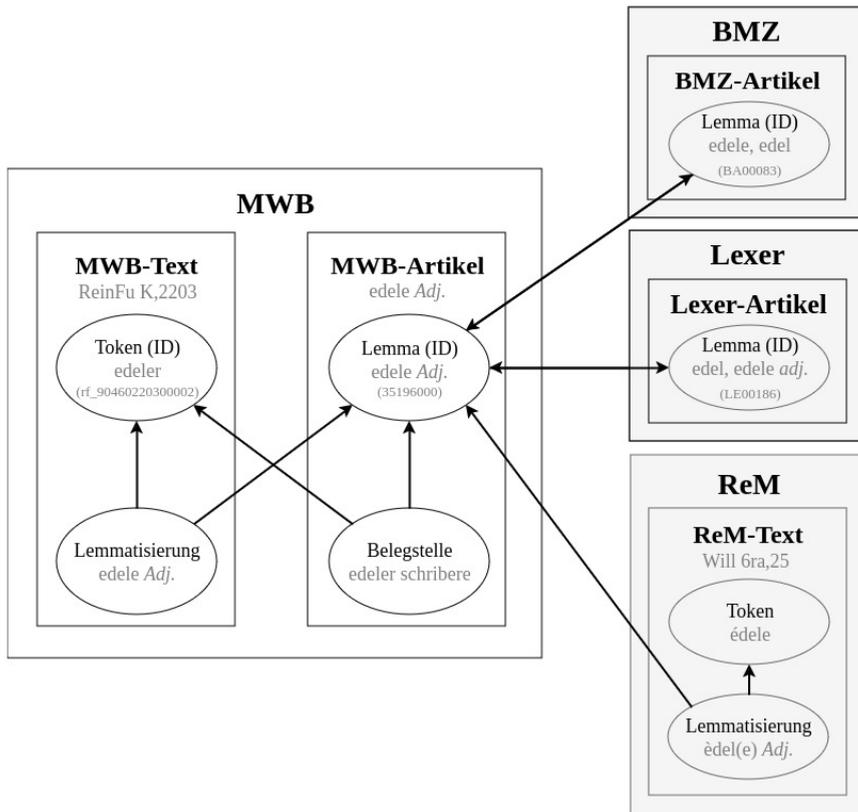
---

<sup>7</sup> Vgl. Nellmann 1997.

<sup>8</sup> Vgl. Fournier 2000.

<sup>9</sup> Vgl. Diehl u. Hansen 2016.

<sup>10</sup> Vgl. Gärtner u. Plate 2013.



**Abb. 1** | Beispielhafte Verlinkung zwischen MWB, BMZ, Lexer und ReM. Graphik: Patrick D. Brookshire und Jonas Richter.

liegen dabei als TEI Lex-0 vor, einem Substandard für lexikalische Ressourcen.<sup>11</sup> Die Primärquellen im MWB-Textarchiv folgen hingegen dem generischeren TEI-P5-Format, wobei streng zwischen dem tokenisierten Volltext mit vor allem typographischen Auszeichnungen und *Stand-off*-Annotationen für Lemmatisierungen unterschieden wird, sodass sich beide Ebenen unabhängig voneinander betrachten lassen. Die Nutzung sogenannter *Feature Structures* ermöglicht eine potentielle (spätere) Anreicherung mit beliebigen weiteren Annotationslayers.<sup>12</sup> Hier können künftige korpuslinguistische Projekte ansetzen, die Flexions- oder Dependenzinformationen eingliedern, oder literaturwissenschaftliche Projekte mit narratologischen Annotationen. Die in der Frühphase des Projekts geschaffene Basis

<sup>11</sup> Vgl. Tasovac, Romary, Banski u. a. 2018.

<sup>12</sup> Vgl. Griesinger u. Stolz 2019, S. 123 f.

eines lemmatisierten Belegarchivs wird im Rahmen der Wörterbucharbeit laufend ergänzt, indem bedarfsweise weitere Textstellen mit den Lemmata über IDs verknüpft werden (vgl. Abb. 1). Automatische Verfahren wie der TreeTagger<sup>13</sup> mit einer mittelhochdeutschen Parameterdatei<sup>14</sup> ergänzen die manuellen Lemmatisierungen, auch wenn dieses Verfahren Lemmaformen der ‚Mittelhochdeutschen Begriffsdatenbank‘ (MHDBDB) ohne die im MWB verwendeten grammatischen Angaben erzeugt. Darüber hinaus wurden neuere Tools wie der RNN-Tagger<sup>15</sup> bereits angetestet. Die *TEI Feature Structures* erlauben eine entsprechende Unsicherheitsmarkierung, sodass in den Daten selbst festgehalten wird, dass im MWB zwar jedes Token lemmatisiert ist, aber nicht alle Lemmatisierungen manuell geprüft sind.

### 3 Schnittstellen

Das Netz mittelhochdeutscher Wörter und Projekte zeichnet sich seit der Jahrtausendwende durch Schnittstellen wie die Mittelhochdeutsche Gesamlemmaliste aus, mit der IDs zwischen Projekten ausgetauscht werden können. Darüber hinaus ist es über die föderierte Architektur des NFDI-Konsortiums Text+ seit dem Jahr 2023 möglich, mittelhochdeutsche Daten über Projektgrenzen hinaus gemeinsam abzufragen. Mit diesen beiden Schnittstellen befassen sich exemplarisch die folgenden Unterabschnitte.

#### 3.1 Mittelhochdeutsche Gesamlemmaliste

In der Vorbereitungsphase des MWB (1994–1999) wurde aus den Einträgen der Vorgängerwörterbücher (BMZ, LEXER samt Nachträgen und Findebuch) eine Lemmaliste kompiliert, die als zentrale Referenz von Anfang an eine Art Rückgrat des Projekts gebildet hat.<sup>16</sup> Einander entsprechende Lemmata aus den älteren Wörterbüchern wurden dabei zusammengeführt. Das Resultat wurde für die Zwecke des MWB bearbeitet und stellt daher kein Gesamtverzeichnis aller Einträge des Vorgängerwörterbuchs dar, da diese unter anderem auch Eigennamen und Verweisformen enthielten. In dem resultierenden Lemmagerüst hat jedes Stichwort eine feste ID und ist mit den zugehörigen Einträgen in den Vorgängerwörterbüchern sowie den nach und nach im MWB entstehenden Artikeln verlinkt. Lemmatisierungen im Volltextarchiv sind

---

13 Vgl. Schmid 1994.

14 Vgl. Echelmeyer, Reiter u. Schulz 2017.

15 Vgl. Schmid 2019.

16 Vgl. Plate u. Recker 2001, S. 177–181.

ebenfalls mit den Stichwörtern in dieser Gesamtlemmaliste verknüpft, die dadurch hervorragend als Rechercheinstrument für Fragestellungen zur mittelhochdeutschen Lexikographie und darüber hinaus taugt. So empfiehlt auch HINKELMANNs, die IDs der MWB-Lemmaliste, auf die auch das ‚Referenzkorpus Mittelhochdeutsch‘ (ReM) verweist, zur eindeutigen Identifizierung mittelhochdeutscher Lemmaformen zu nutzen.<sup>17</sup> Wie sich daraus ein Netz mittelhochdeutscher Wörter entwickelt hat, illustriert Abbildung 1. Sie zeigt, dass im MWB-Textarchiv allen Wortformen (Tokens) eindeutige IDs zugewiesen sind wie zum Beispiel *edeler* in *Saget, edeler schribere* (ReinFu K, 2203).<sup>18</sup> Diese ID wird dann über die *Stand-off*-Lemmatisierung mit der ID des MWB-Lemmas *edele* verlinkt, das über die Lemmaliste bidirektional mit den Äquivalenten *edele*, *edel* im BMZ bzw. *edel*, *edele* im LEXER verknüpft ist. Zusätzlich ist diese Textstelle auch ein Beleg im MWB-Artikel, sodass hier ebenfalls eine projektinterne bidirektionale Verlinkung vorliegt. Diese Vernetzung ist jedoch nicht auf Wörterbuchartikel beschränkt, wie beispielsweise das Token *édele* in *édele uuîntrûbo uóne cýproie* (Will 6ra, 25) aus dem ReM zeigt. Insbesondere Projekte wie das ReM, die ihre Annotationen nicht alphabetisch vornehmen, profitieren dabei von der Tatsache, dass mit der Mittelhochdeutschen Gesamtlemmaliste das MWB auch Verweisziele für noch nicht publizierte Artikel bereithält, um so das Netz der mittelhochdeutschen Ressourcen weiter vergrößern zu können.

### 3.2 Föderierte Suche

An dem von HINKELMANNs beschriebenen Befund, dass gerade lexikalische Ressourcen sich durch eine Vielzahl unterschiedlicher Datenformate auszeichnen,<sup>19</sup> hat sich in den letzten Jahren nur bedingt etwas geändert. Umso wichtiger sind daher Bestrebungen, standardisierte Schnittstellen zu implementieren, die noch mehr Vernetzungen zwischen Projekten ermöglichen. Das MWB hat sich dabei wie das ‚Wörterbuchnetz‘ zunächst im Rahmen des Projektes ‚eHumanities-Zentrum für historische Lexikographie‘ (ZHistLex)<sup>20</sup> und jüngst durch Anbindung an die Föderierte Suche (FCS, *Federated Content Search*) von Text+<sup>21</sup> engagiert. Letztere eröffnet neue Möglichkeiten, da hiermit nicht nur lexikalische Ressourcen wie das MWB und seine Vorgängerwörterbücher, sondern auch die Quellentexte gemeinsam durchsucht werden können. Das kommt der Vision von GRIESINGER und

<sup>17</sup> Vgl. Hinkelmanns 2019, S. 122.

<sup>18</sup> Siglen folgen der MWB-Form – vgl. <http://mhdwb-online.de/quellenverzeichnis.php> (Zugriff: 11.03.2025)

<sup>19</sup> Vgl. Hinkelmanns 2019, S. 133.

<sup>20</sup> Vgl. ebd., S. 140 f., und Pietsch 2020 sowie <https://zhistlex.de/> (Zugriff: 11.03.2025).

<sup>21</sup> Vgl. Brookshire u. Recker-Hamm 2023.

STOLZ, digitale Editionen und lexikalische Ressourcen unter Beachtung gemeinsamer Standards enger miteinander zu verbinden, schon recht nahe.<sup>22</sup>

Die FCS basiert auf standardisierten Abfragesprachen (OASIS-CQL<sup>23</sup> für einfache Suchen, FCS-QL<sup>24</sup> für Korpusuchen und LexCQL<sup>25</sup> für lexikalische Suchen) sowie standardisierten Ergebnisformaten<sup>26</sup> und Fehlerdiagnostiken. Ein projektspezifischer Endpunkt wie der des MWB-Textarchivs übersetzt dabei eine Suchanfrage gemäß der FCS-Syntax in eine interne (XQuery-)Syntax und transformiert die gefundenen (TEI-)Fragmente in das gemeinsame FCS-Format. Dadurch abstrahiert die gemeinsame Suche sowohl von der Datenhaltung als auch dem jeweiligen technischen Setup der Einzelprojekte. So findet zum Beispiel eine Suche nach ‚Terramer‘ über den FCS-Aggregator von Text+<sup>27</sup> nicht nur Treffer in mittelhochdeutschen Projekten wie BMZ, MWB und ReM, sondern unter anderem auch im ‚Deutschen Zeitungsportal der Deutschen Digitalen Bibliothek‘, da die Ausgabe vom 4.10.1941 der ‚Kölnischen Zeitung‘ einen Auszug aus Wolframs von Eschenbach ‚Willehalm‘ enthält. Dieser Zufallsfund könnte für an der Rezeptionsgeschichte Interessierte eine wichtige Rolle spielen, zumal die entsprechende Quelle im Aggregator verlinkt ist. Besteht hingegen Interesse an der Person, kann mit derselben Suche der BMZ-Artikel gefunden werden, aus dem hervorgeht, dass er ‚[S]ohn des Kanabêus‘ war,<sup>28</sup> während die Treffer im MWB und ReM zu weiteren Belegstellen führen. Einer dieser MWB-Belege stammt aus WhvÖst (7724f.) und wird im Aggregator bei einer Multi-Layer-Korpusuche im sogenannten AdvancedDataView folgendermaßen angezeigt:

**Tabelle 1** | Beispiel für FCS-AdvancedDataView-Layer.<sup>29</sup>

Layer	...	s8	s9	s10	s11	s12	s13	s14	s15	...
text	...	der	künc	Terramer/	gewünne	nie	so	manigen	man./	...
lemma	...	der	künic	Terramer	ge-winnen	nie	sô	manigen	man	...
pos	...	DET	NOUN	PROPN	VERB	ADV	ADV	ADJ	NOUN	...

22 Vgl. Griesinger u. Stolz 2019, S. 118, 121.

23 <http://docs.oasis-open.org/search-ws/searchRetrieve/v1.0/os/part5-cql/searchRetrieve-v1.0-os-part5-cql.html> (Zugriff: 11.03.2025).

24 Vgl. van Uytvanck, Olsson, Schonefeld u. a. 2023, S. 39–42.

25 Vgl. Körner, Eckart, Herold u. a. 2023, S. 5f.

26 Vgl. van Uytvanck, Olsson, Schonefeld u. a. 2023, S. 18–25.

27 <https://fcs.text-plus.org/?&query=Terramer&queryType=cql> (Zugriff: 11.03.2025).

28 <https://woerterbuchnetz.de/?sigle=BMZ&lemid=T00393&hitidlist=2705807> (Zugriff: 11.03.2025).

29 <https://fcs.text-plus.org/?&query=%5B%20word%20%3D%20%22Terramer%22%20%5D&queryType=fcs> (Zugriff: 11.03.2025).

In Tabelle 1 sieht man, dass die Korpusuche mit Layern arbeitet und sich auf diese Weise nicht nur durch Anbinden weiterer Projekte horizontal erweitern lässt, sondern auch vertikal durch weitere Annotationslayer. So ließen sich potentiell auch Angaben zur im MWB ausgezeichneten (Un-)Sicherheit der Lemmatisierungen integrieren, um die Nachnutzbarkeit zu erhöhen. Des Weiteren bieten sich auch die für das Netz mittelhochdeutscher Lemmata so zentralen IDs künftig als ein weiterer Layer an. Allerdings kann umgekehrt die Einführung eigener projektspezifischer Layer die Interoperabilität reduzieren, weshalb beide bislang noch nicht eingebunden wurden.

## 4 Nachnutzungsperspektiven

Das MWB kann Anknüpfungspunkt für eine Reihe anderer digitaler mediävistischer Forschungsprojekte sein,<sup>30</sup> wodurch sich wiederum Möglichkeiten für weitere Vernetzungen ergeben. Aus diesem Grund werden in den folgenden Teilabschnitten Nachnutzungsperspektiven der MWB-Daten aufgezeigt, die über die reine Wörterbucharbeit als Hilfestellung zum Verständnis mittelhochdeutscher Texte und die Vernetzung der Daten über IDs hinausgehen. So liegen etwa, wie bereits erwähnt, alle zentralen Ressourcen des MWB als TEI-XML vor, wodurch sie per se nachnutzbar und interoperabel sind.<sup>31</sup> Die automatischen Lemmatisierungen der Primärquellen ermöglichen darüber hinaus statistische Analysen, die einen hohen *Recall* einfordern. Projektintern stehen den Redakteuren bereits Möglichkeiten zur Gruppierung von Belegen nach Textmetadaten zur Verfügung, die gerade bei großen Treffermengen die Sichtung erleichtern, womit sich Abschnitt 4.1 befasst. Hier könnten in Zukunft auch Tasks aus dem Bereich *Natural Language Processing* wie eine Genre-Identifikation oder *Topic Modelling* ansetzen. Mit den manuellen Annotationen lassen sich *Deep Learning*-Modelle umgekehrt durch ihre hohe *Precision* nicht nur zur Lemmatisierung, sondern zum Beispiel auch zur Eigennamenerkennung (nach-)trainieren und/oder evaluieren, wozu im Abschnitt 4.2 eine Pilotstudie präsentiert wird.

### 4.1 Kookkurrenzanalysen

Ein Tool zur Suche nach Kookkurrenzen, das demnächst auch öffentlich verfügbar sein wird, erlaubt die Suche nach Wortformen oder Lemmata in Kombination mit Bedingungen für den Kontext (zum Beispiel ein bestimmtes Wort oder eine

---

<sup>30</sup> Zur Forschungslandschaft vgl. z. B. Bleier, Fischer, Hiltmann u. a. 2019.

<sup>31</sup> Siehe oben unter 2.

Wortart im rechten Kontext). Dabei ist es möglich, alle Lemmatisierungen auszuwerten oder die Suche auf die nicht-automatischen, lexikographisch geprüften Lemmatisierungen zu beschränken.

So zeigt die Umgebungssuche beispielsweise, dass unter 1.504 Treffern zum Adjektiv *reine* häufig Kookkurrenzen mit den Substantiven *maget* (140), *wip* (111) und *got* (135) vorkommen. Befunde, die die Wörterbücher nahelegen oder die der eigenen Erwartungshaltung entsprechen, werden in solchen Suchergebnissen intersubjektiv überprüfbar. Andererseits können auch weniger erwartete Kookkurrenzen wie *reine man* (150) gefunden werden. Der Großteil der Belege hierfür stammt aus dem ‚Märterbuch‘ und Rudolfs von Ems ‚Weltchronik‘, in denen mit dem Ausdruck überwiegend Heilige und biblische Gestalten bezeichnet werden. Bei der Artikelredaktion helfen solche Informationen, um verbreitete Verwendungstypen zu finden. Auch lassen sich gezielt Textstellen suchen, in denen *reine* zum Beispiel in der Nähe von Wortformen steht, die im *Part-of-Speech* (PoS) *Tagging* als Adjektive markiert sind, um zu prüfen, ob es auffallende Verbindungen wie mögliche (Teil-)Synonyme, Antonyme oder Paarformeln gibt.

Das Tool ermöglicht zudem Suchanfragen, die über lexikographische Interessen hinausgehen. So könnte das Textarchiv beispielsweise nach Vorkommen bestimmter Phraseologismen durchsucht oder die Verwendung eines Wortes wie *würfel* (oder semantisch verwandter Wörter) in der Heldenepik unter die Lupe genommen werden, um eine Materialbasis für literaturgeschichtliche oder sprachwissenschaftliche Studien aufzubauen.

Die Suchergebnisse lassen sich auch nach weiteren Kriterien sortieren. Die dafür nötigen Annotationen zur Entstehungsdatierung oder schreibsprachlichen Einordnung der Quellentexte entstanden im Zusammenhang mit dem ZHistLex-Projekt und bedürfen weiterer Ausarbeitung. Ähnliches gilt für die Textsorten-Annotation der rund 1.800 Primärquellen im MWB-Quellenverzeichnis: Diese Daten stammen zu großen Teilen noch aus der Anfangsphase des MWB. Um sie zur Gruppierung oder Filterung der Texte beispielsweise in den Ergebnissen der Umgebungssuche verwenden zu können, ist eine Vereinheitlichung und ontologische Aufbereitung notwendig. Dafür stehen im Rahmen der Wörterbucharbeit jedoch weder Zeit noch Mittel zur Verfügung, künftige Projekte könnten aber diese Anlagen weiter ausbauen. So wäre etwa die Verknüpfung der MWB-Quellen mit der Textreihentypologie von ZEPPEZAUER-WACHAUER und HEILES, einem SKOS-Vokabular,<sup>32</sup> wünschenswert. Der Nutzen solcher Datenanreicherung ist offensichtlich, würde er doch nicht nur eine Filterung des Primärquellenverzeichnisses ermöglichen, sondern auch anhand der Quellenzitate im Wörterbuch eine Auswertung des Wortschatzes nach diesen Metadaten erlauben.

---

<sup>32</sup> Vgl. Zepezauer-Wachauer u. Heiles 2023. SKOS (*Simple Knowledge Organization System*) ist ein vom W3C empfohlener Standard zur Kodierung kontrollierter Vokabulare.

Hier zeigt sich ein generelles Problem von Langzeitprojekten, die in ihrer Laufzeit zwar an einer umfangreichen Aufgabe sehr intensiv arbeiten können, andererseits aber an einen Projektrahmen gebunden sind, der über viele Jahre – oft Jahrzehnte – bestehen bleibt und nicht flexibel auf neue Anforderungen reagiert, die sich in der Fachwelt etablieren. Dies hat zur Folge, dass Forschungsdaten im Rahmen eines vor langer Zeit gestarteten Projekts – wenn überhaupt – nur in der Form eines wenig befriedigenden Kompromisses an konzeptuelle oder technische Gegebenheiten (etwa Datenstandards), die heutzutage als offensichtliches Desiderat gelten mögen, angepasst werden können.

## 4.2 Eigennamenerkennung

Während Kookkurrenzanalysen ein Beispiel für eine stärker korpuslinguistisch motivierte Nachnutzungsperspektive der MWB-Daten sind, bildet eine automatische Erkennung von Eigennamen (*Named Entity Recognition*, NER) Anknüpfungspunkte für Projekte aus den Bereichen *Digital Humanities* und *Computational Literary Studies*. Ein Beispiel hierfür sind die (Figuren-)Netzwerkanalysen von BRAUN und KETSCHIK, die auf der Auswertung von Figurennennungen basieren.<sup>33</sup> „Personendaten“ sind mit den Worten von HINKELMANNs „natürlich nicht das Kerngeschäft des MWBs oder der MHDBDB“.<sup>34</sup> Nichtsdestotrotz wurden in einigen Texten Personen- und Ortsnamen manuell ausgezeichnet, wenn auch ohne Differenzierung zwischen diesen beiden Entitätentypen. Diese Annotationen lassen sich als Basis für NER-Algorithmen nachnutzen, um in beliebigen unausgezeichneten mittelhochdeutschen Texten entsprechende Wortformen zu finden. Eine Möglichkeit ist dabei ein simples *String Matching* der erfassten Wortformen (Types), eine weitere das Finetuning eines vortrainierten *Deep Learning*-Modells mit den Annotationen. Für eine diesbezügliche Pilotstudie wurden 3.000 zufällige Textstellen aus elf annotierten Texten ausgewählt und diese in Trainings- bzw. Testdaten aufgeteilt.

---

<sup>33</sup> Vgl. Braun u. Ketschik 2019.

<sup>34</sup> Hinkelmanns 2019, S. 139.

**Tabelle 2** | Anzahl der Textstellen im Trainings- und Testdatenset pro Text (eingeklammert die Anzahl der Textstellen, die mindestens eine Entität enthalten).

Text	Trainings-Textstellen	Test-Textstellen	Summe
Parz	1.032 (724)	263 (185)	1.295 (909)
Wh	625 (473)	150 (117)	775 (590)
UvZLanz	267 (174)	68 (41)	335 (215)
RvEBarl	141 (117)	42 (34)	183 (151)
Er	107 (83)	18 (16)	125 (99)
Wig	104 (97)	20 (17)	124 (114)
VAlex	90 (68)	29 (23)	119 (91)
Vateruns	25 (5)	9 (4)	34 (9)
VEzzo	4 (1)	1 (1)	5 (2)
SuTheol	2 (2)	0 (0)	2 (2)
SüklZw	1 (1)	0 (0)	1 (1)
Summe	2.400 (1746)	600 (438)	3.000 (2184)

Die Textstellen sind hier wie beim Tool zur Kookkurrenzanalyse *Keywords in Context* mit je zehn vorangehenden bzw. nachfolgenden Tokens um eine beliebige großgeschriebene Wortform (vgl. Tab. 3), wodurch sie strukturell Ergebnissen aus der FCS ähneln.<sup>35</sup> Als Keywords wurden dabei bewusst nicht nur Eigennamen gewählt, damit das Sprachmodell nicht einfach nur die Features Großschreibung oder Position in der Textstelle als diskriminierendes Merkmal lernen kann. Letzteres wird zudem durch die Tatsache erschwert, dass im Kontextfenster weitere Eigennamen vorkommen können wie beispielsweise *Pantanor* im linken Kontext in Tabelle 3. Insgesamt umfasst das Datenset 62.916 Tokens, wovon 4.467 (~7,1 Prozent) manuell als Eigenname ausgezeichnet sind.

---

<sup>35</sup> Vgl. Tabelle 1 und 3.

**Tabelle 3** | Beispiele für Trainings- bzw. Test-Textstellen (Eigennamen hervorgehoben).

Quelle	linker Kontext					Keyword	rechter Kontext				
<b>Parz 566,1</b>	...	er	in	dienen	wolte.	Da!	wärn	si	doch	unschuldec	...
		○	○	○	○	○	○	○	○	○	
<b>Wh 77,24</b>	...	<b>Pantanor</b>	gap	dem	künege	<b>Salatrê,</b>	der	gabz	dem	künege	...
		<b>B</b>	○	○	○	<b>B</b>	○	○	○	○	

Die Eigennamenannotation der Trainings- und Testdaten folgt dem verbreiteten IOB-Schema, das auf RAMSHAW und MARCUS zurückgeht,<sup>36</sup> wobei nur zwischen den Labeln *O* (= keine Entität) und *B* (= erstes Token einer Entität) unterschieden wird. Das Label *I* (= Folgetoken einer zusammengesetzten Entität) wird nicht benötigt, da in diesen Texten Mehrwort-Entitäten keine nennenswerte Rolle spielen. Durch den Rückgriff auf dieses Standardformat ist eine spätere Einbindung in verschiedene Tools denkbar.

Mit den Trainingsdaten wurde ein Finetuning zweier deutscher Transformermodelle mit dem Python-Modul Transformers<sup>37</sup> sowie Default-Parametern über drei Epochen durchgeführt.<sup>38</sup> Das erste Modell GHistBERT wurde von BECK und KÖLLNER auf ca. 660.000 Sätzen aus Texten vortrainiert, die zwischen 750 und 1999 entstanden sind.<sup>39</sup> In diesen Daten finden sich nach Angaben der Autorinnen auch 246.000 (~37,5 Prozent) mittelhochdeutsche Sätze aus dem ReM, weshalb die Möglichkeit besteht, dass es einige Textstellen bereits gesehen hat. Das ist bei dem anderen in dieser Pilotstudie gewählten Modell ‚gbase-bert‘<sup>40</sup> nicht zu erwarten, weil hier die Trainingsgrundlage 163,4 GB deutschsprachige Internetseiten sind. Um einschätzen zu können, wie gut diese Modelle Entitäten finden können, wurden, wie bei entsprechenden Benchmarks üblich, neben der sogenannten *Majority Baseline*, die überhaupt keine Entitäten erkennt, noch weitere bereits existierende Methoden getestet. Das erste Verfahren ist dabei der TreeTagger von SCHMID<sup>41</sup>, der eigentlich ein PoS-Tagger ist und als solcher im MWB für automatische Lemmatisierungen genutzt wird.<sup>42</sup> Allerdings entspricht der PoS-Tag PROPN im Prinzip der Eigennamenannotation im MWB, sodass diese Methode einem Mapping auf das IOB-Schema

36 Vgl. Ramshaw u. Marcus 1995.

37 Wolf, Debut, Sanh u. a. 2020.

38 Für entsprechende Daten und Skripte siehe <https://github.com/digicademy/mwb-name-recognition> (Zugriff: 11.03.2025)

39 Vgl. Beck u. Köllner 2023.

40 Chan, Schweter u. Möller 2020.

41 Vgl. Schmid 1994; ergänzt um die mittelhochdeutsche Parameterdatei von Echelmeyer, Reiter u. Schulz 2017.

42 Siehe oben unter 2.

vergleichbar ist. Als Letztes wurde mit ‚stanza‘<sup>43</sup> noch ein eigentliches NER-Modell getestet, das auf Zeitungsartikeln von 1992<sup>44</sup> und damit ebenso wie ‚gbert-base‘ nicht auf mittelhochdeutschen Texten basiert. Um die Vergleichbarkeit zu gewährleisten, wurden hier ebenfalls nur Personen und Orte berücksichtigt und weitere Typen auf O (= kein Eigenname) gemappt. Die Klassifikationsergebnisse all dieser Verfahren sowie des *String Matchings* bezogen auf die vier üblichen Kennzahlen *Accuracy*, *Precision*, *Recall* und *makro F1-Score* finden sich in Tabelle 4. Die *Accuracy* misst dabei, wie viel Prozent der Testdatensätze insgesamt richtig klassifiziert wurden, während die anderen Maße sich auf die relevanten Modellvorhersagen beziehen. So misst die *Precision*, wie viel Prozent der vom Modell als Eigenname ausgezeichneten Tokens auch wirklich Eigennamen sind, der *Recall*, wie viel Prozent der Eigennamen gefunden wurden, und der F1-Score kombiniert beide Werte.

**Tabelle 4** | Klassifikationsergebnisse verschiedener Methoden zur Eigennamenerkennung (beste Ergebnisse fett hervorgehoben; bei Modellen mit \* erfolgte ein Finetuning).

Methoden	Datengrundlage	Accuracy	Precision	Recall	F1 (makro)
Majority Baseline	-	93,0%	0,0%	0,0%	48,2%
String Matching	MWB-Wortformenliste	96,6%	<b>99,6%</b>	52,0%	83,3%
TreeTagger	MHDBDB	98,2%	96,2%	77,7%	92,5%
stanza	CoNLL03	94,1%	86,8%	19,3%	64,3%
GHisBERT*	MWB-Textarchiv	97,6%	80,7%	86,3%	91,1%
gbert-base*	MWB-Textarchiv	<b>99,4%</b>	95,6%	<b>95,7%</b>	<b>97,7%</b>

Erwartungsgemäß sind die Werte aller vier Kennzahlen bei der *Majority Baseline* am niedrigsten. Allerdings verdeutlicht hier die *Accuracy* von rund 93,0 Prozent die erwähnte Seltenheit von Eigennamen in diesem Datenset. Bezogen auf *Accuracy* und F1-Score war das ‚stanza‘-NER-Modell für modernes Deutsch am zweitschlechtesten, was primär am niedrigen *Recall* liegt. Grund hierfür dürfte die geringe Anzahl an für mittelhochdeutsche Texte relevanten Eigennamen in heutigen Zeitungsartikeln sein. Einen zwar deutlich besseren, aber ebenfalls im Vergleich niedrigen *Recall* erzielte auch der *String-Matching*-Ansatz, der jedoch wegen seiner mit Abstand höchsten *Precision* von 99,6 Prozent immer noch einen

<sup>43</sup> Qi, Zhang, Zhang u. a. 2020, S. 142 f.

<sup>44</sup> Tjong Kim Sang u. De Meulder 2003.

hohen Nachnutzwert hat. Ähnliches gilt für den TreeTagger, der zwar etwas weniger präzise war, aber dafür einen deutlich höheren *Recall* einbrachte. Die beiden nachtrainierten Modelle waren in Bezug auf das Verhältnis von *Precision* und *Recall* (und damit den F1-Score) ausgewogener. Interessanterweise war beim diachronen Modell jedoch die *Precision* am zweitniedrigsten, während das mit modernem Deutsch vortrainierte ‚gbert-base‘ mit einer *Accuracy* von 99,4 Prozent und einem F1 von 97,7 Prozent insgesamt am besten abschnitt. Möglicherweise liegt das daran, dass dieses Modell besser generalisieren kann und daher beim Finetuning der Transfer besser gelang. Somit könnte dieses neue Modell schon jetzt in anderen Projekten zur automatischen Eigennamenerkennung in mittelhochdeutschen Texten genutzt werden. Darüber hinaus unterstreichen die Ergebnisse, dass, selbst wenn im MWB Annotation und Lemmatisierung von Eigennamen nur tentativ erfolgten, diese Anlage nachgenutzt werden kann – sei es innerhalb des Projektes zum *Pretagging* noch unausgezeichneter Eigennamen oder zur Vorbereitung anderer Personen- bzw. Ortsnamen-basierter Analysen.

## 5 Fazit

Mit der Bitte des Verlegers Salomon HIRZEL, dass Matthias LEXER ein Handwörterbuch verfassen möge, das sich eng auf das Wörterbuch von BENECKE-MÜLLER-ZARNCKE bezieht, begann die Vernetzung des mittelhochdeutschen Wortschatzes. Heute ist das Netz gewachsen und weist weit über seinen ursprünglichen Horizont hinaus. Weitverbreitete Datenformate und Standards zu Schnittstellen ermöglichen, Annotationslayer in Einzeltexten oder Textkorpora mit Verweisen auf Einträge in der Lemmaliste des MWB anzureichern und damit nicht nur Wortformen zu lemmatisieren, sondern sie indirekt auch mit den einschlägigen Artikeln in den mittelhochdeutschen Wörterbüchern zu verbinden. Schnittstellen, die zwischen nativen Datenformaten und externen Anfragen ‚übersetzen‘ können, erlauben der Forschung, mit föderierten Suchen wie der FCS von Text+ gezielte Fragen an eine große Anzahl nicht nur mediävistischer Projekte zu stellen. Nicht immer sind die technischen Grundlagen dafür bei Langzeitprojekten von Anfang an gegeben, wie es beispielsweise bei den IDs der Lemmaliste der Fall ist. In anderen Fällen, wie etwa bei den Schnittstellen, die die Anbindung an die FCS ermöglichen, müssen Drittmittel eingeworben werden. Die Pilotstudie zur Entitätenerkennung lässt aber hoffen, dass auch diejenigen Datenanreicherungen, die nicht mit größter Konsequenz und Einheitlichkeit, sondern eben nur tentativ ‚nebenbei‘ vorgenommen werden, bereits einen Mehrwert darstellen und als Basis für weiterführende Untersuchungen dienen können. Dazu müssen die Vernetzungsmöglichkeiten prinzipiell offen und erweiterbar angelegt sein.

Die am Beispiel mittelhochdeutscher Daten hier exemplarisch vorgestellte Verbesserung der Zugänglichkeit und Interoperabilität steht dabei für eine generelle Tendenz in der Mediävistik, die in Zukunft zu einer noch stärker interdisziplinären Forschung führen kann. Denn gerade durch die Vernetzung unterschiedlicher Informationen können neue Muster identifiziert und analysiert werden.

## Literaturverzeichnis

- Beck, Christin u. Marisa Köllner:** GHISBERT – Training BERT from Scratch for Lexical Semantic Investigations Across Historical German Language Stages. In: Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change. Singapur 2023, S. 33–45. <https://doi.org/10.18653/v1/2023.lchange-1.4> (Zugriff: 11.03.2025).
- Bleier, Roman, Franz Fischer, Torsten Hiltmann u. a. (Hgg.):** Digitale Mediävistik (Das Mittelalter. Perspektiven mediävistischer Forschung 24). Berlin 2019.
- BMZ = Mittelhochdeutsches Wörterbuch, mit Benutzung des Nachlasses von Georg Benecke ausgearb. von Wilhelm Müller und Friedrich Zarncke. Leipzig 1854–66. Digitalisierte Fassung im Wörterbuchnetz des Trier Center for Digital Humanities, Version 01/23, <https://www.woerterbuchnetz.de/BMZ> (Zugriff: 11.03.2025).
- Brookshire, Patrick D. u. Ute Recker-Hamm:** Text+Plus, #03: Verlinkungen und standardisierte Suchen: MWB-APIplus. In: Text+ Blog. 2023. <https://textplus.hypotheses.org/8245> (Zugriff: 11.03.2025).
- Chan, Branden, Stefan Schweter u. Timo Möller:** German’s Next Language Model. In: Proceedings of the 28th International Conference on Computational Linguistics. Barcelona 2020, S. 6788–6796. <https://doi.org/10.18653/v1/2020.coling-main.598> (Zugriff: 11.03.2025).
- Diehl, Gerhard u. Nils Hansen:** Zwischen Handschrift und Online-Datenbank. Bemerkungen zur Lexikographie des Mittelhochdeutschen. In: Anja Lobenstein-Reichmann u. Peter O. Müller (Hgg.): Historische Lexikographie zwischen Tradition und Innovation (Studia Linguistica Germanica 129). Berlin 2016, S. 287–305.
- Echelmeyer, Nora, Nils Reiter u. Sarah Schulz:** Ein PoS-Tagger für „das“ Mittelhochdeutsche. In: Jahrestagung des Verbands Digital Humanities im deutschsprachigen Raum. Bern 2017, S. 141–147. <http://dx.doi.org/10.18419/opus-9023> (Zugriff: 11.03.2025).
- Fournier, Johannes:** Digitale Dialektik: Chancen und Probleme mittelhochdeutscher Wörterbücher in elektronischer Form. In: Herbert Ernst Wiegand (Hg.): Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium (Lexicographica Series Maior 100). Tübingen 2000, S. 85–108.
- Gärtner, Kurt:** Das Handexemplar von Matthias Lexers ‚Mittelhochdeutschem Handwörterbuch‘. In: Horst Brunner (Hg.): Matthias von Lexer. Beiträge zu seinem Leben und Schaffen (Zeitschrift für Dialektologie und Linguistik, Beiheft 80). Stuttgart 1993, S. 109–131.

**Gärtner, Kurt, Christoph Gerhardt,**

**Jürgen Jährling u. a.:** Findebuch zum mittelhochdeutschen Wortschatz. Stuttgart 1992.

**Gärtner, Kurt u. Ralf Plate:** Die Doppelfunktion des digitalen Textarchivs als Wörterbuchbasis und als Komponente der Online-Publikation. Am Beispiel des Mittelhochdeutschen Wörterbuchs. In: Ingelore Hafemann (Hg.): Perspektiven einer corpusbasierten historischen Linguistik und Philologie. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen Akademie der Wissenschaften, 12.–13. Dezember 2011. Berlin 2013, S. 193–204. URN: urn:nbn:de:kobv:b4-opus-24400

**Griesinger, Christian u. Michael Stolz:**

Sprachwissenschaftliche Erschließungsmethoden für digitale Editionen mittelhochdeutscher Texte. In: Bleier, Fischer, Hiltmann u. a. 2019, S. 112–128.

**Hinkelmanns, Peter:** Mittelhochdeutsche Lexikographie und Semantic Web. Die Anbindung der ‚Mittelhochdeutschen Begriffsdatenbank‘ an Linked Open Data. In: Bleier, Fischer, Hiltmann u. a. 2019, S. 129–141.

**Körner, Erik, Thomas Eckart, Axel Herold u. a.:** Federated Content Search for Lexical Resources (LexFCS): Specification. Version 0.1, 2023-05-04. <https://doi.org/10.5281/zenodo.7986303> (Zugriff: 11.03.2025).

**Lexer, Matthias:** Mittelhochdeutsches Handwörterbuch. Zugleich als Supplement und alphabetischer Index zum Mittelhochdeutschen Wörterbuch von Benecke-Müller-Zarncke. Band 1, Stuttgart 1872.

**Nellmann, Eberhard:** Die mittelhochdeutschen Wörterbücher. Ihre Qualitäten, ihre Grenzen, ihre mögliche Erneuerung.

In: Eijiro Iwasaki u. Yoshinori Shichiji (Hgg.): Begegnung mit dem „Fremden“: Grenzen – Traditionen – Vergleiche. Akten des VIII. Internationalen Germanisten-Kongresses, Tokyo 1990. Bd. 4. Sektion 4, Kontrastive Syntax; Sektion 5, Kontrastive Semantik, Lexikologie, Lexikographie; Sektion 6, Kontrastive Pragmatik. München 1991, S. 254–263.

**Nellmann, Eberhard:** Quellenverzeichnis zu den mittelhochdeutschen Wörterbüchern. Ein kommentiertes Register zum ‚Benecke/Müller/Zarncke‘ und zum ‚Lexer‘. Stuttgart, Leipzig 1997.

**Pietsch, Andre:** Historische Wörterbücher und Wörterbuchsysteme in digitalen Umgebungen: Komponenten, Verknüpfungen, Darstellungsmittel, Nutzungsoptionen. In: ZHistLex-Papiere. 2020. [https://zhistlex.de/papiere/pietsch\\_2020\\_digitale\\_woerterbuecher\\_ZHistLex.pdf](https://zhistlex.de/papiere/pietsch_2020_digitale_woerterbuecher_ZHistLex.pdf) (Zugriff: 11.03.2025).

**Plate, Ralf u. Ute Recker:** EDV für Wörterbuchzwecke und neue lexikographische Arbeitsweisen. Erfahrungen beim Aufbau des elektronischen Text- und Belegarchivs für das mittelhochdeutsche Wörterbuch. In: Stephan Moser, Peter Stahl, Werner Wegstein u. a. (Hgg.): Maschinelle Verarbeitung altdeutscher Texte V: Beiträge zum Fünften Internationalen Symposium, Würzburg 4.–6. März 1997. Berlin, New York 2001, S. 169–184. <https://doi.org/10.1515/9783110909494.169> (Zugriff: 11.03.2025).

**Qi, Peng, Yuhao Zhang, Yuhui Zhang u. a.:** Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2020, S. 101–108.

<https://doi.org/10.18653/v1/2020.acl-demos.14> (Zugriff: 11.03.2025).

**Ramshaw, Lance A. u. Mitchell P. Marcus:**

Text Chunking Using Transformation-Based Learning. In: Proceedings of the Third ACL Workshop on Very Large Corpora. Cambridge 1995, S. 82–94. <https://aclanthology.org/W95-0107/> (Zugriff: 11.03.2025).

**Schmid, Helmut:** Probabilistic Part-of-Speech Tagging Using Decision Trees.

In: Proceedings of the International Conference on New Methods in Language Processing. Manchester 1994, S. 44–49. <https://cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf> (Zugriff: 11.03.2025).

**Schmid, Helmut:** Deep Learning-Based

Morphological Taggers and Lemmatizers for Annotating Historical Texts. In: Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage. Brüssel 2019, S. 133–137. <https://doi.org/10.1145/3322905.3322915> (Zugriff: 11.03.2025).

**Tasovac, Toma, Laurent Romary, Piotr**

**Banski u. a.:** TEI Lex-0: A Baseline Encoding for Lexicographic Data. Version 0.9.3. DARIAH Working Group on Lexical Resources. 2018. <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html> (Zugriff: 11.03.2025).

**Tjong Kim Sang, Erik F. u. Fien De Meulder:**

Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In: Proceedings of the seventh conference on Natural

language learning at HLT-NAACL 4. Edmonton 2003, S. 142–147.

<https://doi.org/10.3115/1119176.1119195> (Zugriff: 11.03.2025).

**Uytvanck, Dieter van, Leif-Jöran Olsson,**

**Oliver Schonefeld u. a.:** CLARIN Federated Content Search (CLARIN-FCS) – Core 2.0. Version 1.0, 2023-04-26. <https://office.clarin.eu/v/CE-2017-1046-FCS-Specification-v20230426.pdf> (Zugriff: 11.03.2025).

**Wilkinson, Mark D., Michel Dumontier,**

**IJsbrand Jan Aalbersberg u. a.:** The FAIR Guiding Principles for Scientific Data Management and Stewardship. In: Scientific Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18> (Zugriff: 11.03.2025).

**Wolf, Thomas, Lysandre Debut, Victor**

**Sanh u. a.:** Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2020, S. 38–45. <https://doi.org/10.18653/v1/2020.emnlp-demos.6> (Zugriff: 11.03.2025).

**Zeppetzauer-Wachauer, Katharina u.**

**Marco Heiles:** Eine digitale Textreihentypologie für deutschsprachige Texte des Mittelalters und der Frühen Neuzeit. Showcase eines kontrollierten Vokabulars in SKOS. In: Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte 6 (2023), S. 6–39. <https://doi.org/10.26012/mittelalter-30680> (Zugriff: 11.03.2025).

# Linked Open Bibliographical Data am Beispiel der internationalen Artusforschung

## Wie Literaturangaben mehr leisten, als Literatur anzugeben

### Kontakt

#### Abdallah Mohamed

##### Abdallah Abdelnaby,

Georg-August-Universität Göttingen,  
Niedersächsische Staats- und  
Universitätsbibliothek,  
Platz der Göttinger Sieben 1,  
37073 Göttingen,  
abdelnaby@sub.uni-goettingen.de  
 <https://orcid.org/0009-0006-1243-1854>

#### Prof. Dr. Nathanael Busch,

Philipps-Universität Marburg,  
Deutsche Philologie des Mittelalters,  
Deutschausstraße 15,  
35032 Marburg,  
buschn@uni-marburg.de  
 <https://orcid.org/0000-0001-9683-113X>

#### Dr. Julika Moos,

Philipps-Universität Marburg,  
Deutsche Philologie des Mittelalters,  
Deutschausstraße 15,  
35032 Marburg,  
julika.moos@uni-marburg.de  
 <https://orcid.org/0009-0006-0958-9781>

#### Dr. Lilja Mareike Sautter,

Georg-August-Universität Göttingen,  
Niedersächsische Staats- und  
Universitätsbibliothek,  
Platz der Göttinger Sieben 1,  
37073 Göttingen,  
sautter@sub.uni-goettingen.de,  
 <https://orcid.org/0000-0003-3898-1134>

**Abstract** Research bibliographies are not just a tool for finding information in times of high publication numbers – they also represent research data that can be linked to other data. In this article, we examine how bibliographies can be used as linked open data (LOD) for future research and teaching. Our aim is to show under which conditions a bibliographic database can provide new access by linking structured data, annotated information, and suitable contexts. This paper discusses these conditions using the DFG pilot project ‘Artus digital’ as an example. This project is currently creating an open knowledge base for international medieval research. The project’s merging of analogue bibliographies with a digital database will not only connect research bases of the 20th and 21st centuries but also become a long-term contribution to the digital infrastructure of Arthurian research. It will preserve older knowledge, prepare it in a future-orientated manner, continuously update it, and thereby create guidelines for the further digitalisation of bibliographic data.

**Keywords** Arthurian Scholarship; Bibliography; Digital Medieval Studies; Linked Open Data; OCR

## 1 Suchen und finden

Wer schreiben will, muss lesen. Erst aus den Widersprüchen der vorgefundenen Literatur ergibt sich, was zu schreiben ist. Die Suche nach anregender Literatur ist eine wichtige Bedingung für das Gelingen des bevorstehenden Schreibprozesses. Geisteswissenschaftliches Arbeiten verdankt sich oft Lektürezufällen, die aller Zufälligkeit zum Trotz gut vorbereitet sein müssen: Solche intendierten Zufälle<sup>1</sup> setzen nicht nur individuelle Erfahrung voraus, sondern auch gut organisierte Institutionen, die die einschlägige Literatur zur Verfügung stellen. Die Bibliothekslandschaft mag sich in den letzten Jahren verändert haben, doch das Stöbern in thematisch sortierten Regalen einer (Fachbereichs-)Bibliothek bleibt eine beständige Strategie vieler Wissenschaftler\*innen, damit sie ‚zufällig‘ mehr finden als das, was eigentlich gesucht wurde. Oft reicht für sie nur die Signatur eines einzelnen Bandes, um vor einem thematisch sortierten Bücherregal in eine neue Forschungswelt einzutauchen, weil sich links und rechts von diesem Buch nützliche Titel finden lassen, auf die sie in ihren Recherchen noch nicht gestoßen sind. Voraussetzung ist allerdings, dass das Bücherregal vor der Recherche nach institutionellen Entscheidungen sortiert worden ist, die den Suchenden später einen Weg passender Zufälle vorgeben.

Doch die Bibliotheksbenutzung ist im Umbruch – und in ihrer Folge auch die Bibliothekslandschaft: Regalbestände werden reduziert, Institutsbibliotheken aufgelöst, und zahlreiche Publikationen sind inzwischen ausschließlich online verfügbar. Dieser Umbruch schränkt die hergebrachte Strategie, verwandte Titel zu finden, ein, aber es sind bereits neue Strategien entwickelt worden: Bekanntlich schlagen Online-Verkaufshäuser, Webseiten von Verlagen und zuweilen sogar Bibliotheken weitere Titel vor mit dem Hinweis „andere Kund\*innen kauften auch“ oder „das könnte Sie auch interessieren“. Solche Empfehlungen – man nennt sie im Technikjargon auch ‚Assoziationsrelationen‘ – haben durchaus das Potential, ‚Zufallsfunde‘ abseits der inhaltsbezogenen Suchanfragen zu liefern. Allein die Algorithmen, die just diese Vorschläge unterbreiten, bleiben undurchschaubar und sind keineswegs mit der Zuverlässigkeit einer bibliothekarischen Expertise vergleichbar. Wissenschaftliche Suchstrategien sollten nicht bloß das zutage fördern, was bereits massenhaft gelesen wurde, sondern auch weniger offensichtliche Zufallsfunde ermöglichen. Andernfalls wird im schlimmsten Fall am Ende nicht das gelesen, was relevant ist, sondern was andere gekauft haben.

Es gibt jedoch Grund zur Zuversicht: Der Umbruch eröffnet auch neue Informationszugänge, die den technischen Entwicklungen und den veränderten

---

1 Vgl. Kennedy, Whitehead u. Ferdinand-James 2022. Der Recherchefaktor Serendipität hat vor allem im Bereich der *Library and Information Sciences* (LIS) eine lange Forschungsgeschichte, dazu etwa: Erdelez, Heinström, Makri u. a. 2016.

Bedürfnissen der Suchenden Rechnung tragen. Anstatt in Nostalgie zu verfallen, lohnt es sich zu fragen, wie sich die inspirierende Erfahrung, an einem Bibliotheksregal thematisch angeordnete Bände durchzusehen, auf digitale Medien übertragen lässt. Was ist zu beachten? Eine mögliche Antwort lässt sich im Konzept der *Linked Open Data* (LOD) finden, dessen ambitioniertes Ziel es ist, Verbindungen zu kreieren, die von den einzelnen Datenanbietern nicht vorhergesehen werden.<sup>2</sup> In diesem Beitrag möchten wir zeigen, inwiefern es von Nutzen ist, bibliographische Daten, die oftmals als reines Mittel zum Zweck behandelt werden oder gar als ‚Abfallprodukt‘ wissenschaftlicher Tätigkeit entstehen, als offene Forschungsdaten ernst zu nehmen, und welche Schritte erforderlich sind, damit sie als LOD für wissenschaftliche Recherchen fruchtbar gemacht werden können.

Um damit zu beginnen, wie wissenschaftliche Recherche in der Zukunft aussehen könnte, ein Beispiel: Angenommen, man interessiert sich für die Ortsnamen in Wolframs von Eschenbachs Roman ‚Parzival‘, weil man die darin erwähnten Burgen in Franken erforschen möchte. Wenn man in einem Online-Bibliothekskatalog (OPAC) nach dem Titel und dem Namen des Verfassers sucht, würde man das Ergebnis noch mit einem relevanten Schlagwort weiter eingrenzen. Die (hypothetische) Katalogrecherche liefert zum einen Literaturhinweise auf germanistisch-mediävistische Forschungsbeiträge zur Ortsnamenkunde des ‚Parzival‘ und zum anderen auf die ‚Mittelhochdeutsche Begriffsdatenbank‘, in der alle Ortsnamen ausgegeben werden, die im Kontext des Wortes *bure* erscheinen.<sup>3</sup> Von dort gelangt man direkt zum ‚Handschriftencensus‘, der es ermöglicht, die Textzeugen nach ihrer sprachlichen Herkunft zu sortieren, so dass man sich sämtliche Handschriften aus dem oberdeutschen Sprachraum auflisten lassen kann.<sup>4</sup> Ein weiterer Klick führt zum ‚Parzival-Projekt‘, das Transkriptionen und Abbildungen jener Belegstellen dieser Handschriften präsentiert.<sup>5</sup> Mit diesen Ortsnamen ausgestattet, wird man im Bibliothekskatalog auf die Ortsdatenbank der Bayerischen Staatsbibliothek<sup>6</sup> stoßen, die wiederum durch das Portal ‚bavarikon‘<sup>7</sup> gespeist wird, das bayerische Kunst-, Kultur- und Wissensgüter verzeichnet. Über dieses Portal erkundet man reale Orte im süddeutschen Raum, die im ‚Parzival‘

---

2 Vgl. Borek 2023, S. 90.

3 ‚Mittelhochdeutsche Begriffsdatenbank‘ (MHDBDB), <https://mhdbdb.sbg.ac.at/#/> (Zugriff: 13.03.2025).

4 ‚Handschriftencensus‘: Wolfram von Eschenbach: ‚Parzival‘, <https://handschriftencensus.de/werke/437> (Zugriff: 13.03.2025).

5 ‚Parzival-Projekt. Ein Projekt des Schweizerischen Nationalfonds und der Deutschen Forschungsgemeinschaft‘, <https://parzival.unibe.ch> (Zugriff: 13.03.2025).

6 ‚Ortsdatenbank‘, <https://www.bsb-muenchen.de/sammlungen/bavarica/recherche/ortsdatenbank/> (Zugriff: 13.03.2025).

7 ‚Bavarikon‘, <https://www.bavarikon.de> (Zugriff: 13.03.2025).

erwähnt werden – wie etwa Abenberg oder Dollnstein –, denn im Portal sind Informationsquellen wie Digitalisate historischer Karten und Schriften direkt verlinkt. Mit nur wenigen Klicks lassen sich bei der Recherche geschichts- und literaturwissenschaftliche Forschungsansätze miteinander verknüpfen, was dazu einlädt, innovative interdisziplinäre Fragestellungen zu entwickeln. Will man dann im Kontext realer Orte mehr über die Burg Wildenberg im Odenwald erfahren, die als Entstehungsort des ‚Parzival‘ gehandelt wird, bietet das Ergebnis der Recherche eine weitere Verknüpfung zu einer Datenbank an: In der ‚Castles Inventory Database‘ des Europäischen Burgeninstituts, kurz EBIDAT,<sup>8</sup> werden historische Informationen, Fotos, Karten und Forschungsliteratur zu dieser Burg aufbereitet. In kürzester Zeit lässt sich ein Eindruck der Burg im Mittelalter, ihrer lokalen Verortung im nördlichen Bayern und ihres heutigen ruinösen Zustands verschaffen und sogar eine geeignete Verkehrsrouten dorthin vorschlagen.

Falls bei der Recherche nicht die großen historischen Artefakte, sondern Bildzeugnisse des ‚Parzival‘-Stoffes interessieren, sind auch diese ohne größere Umstände recherchierbar. Denn der OPAC verzeichnet nicht nur einschlägige wissenschaftliche Studien aus dem Bereich der Text-Bild-Forschung, sondern leitet auch zu relevanten Bilddatenbanken weiter. Beispielsweise wird auf das ‚Gothic Ivories Project‘ des Courtauld Institute of Art<sup>9</sup> verwiesen, in dem direkt die Elfenbeinkästchen gefunden werden, die mit dem berühmten arthurischen Stoff in Verbindung gebracht werden. Mit nur wenigen Klicks ist das Elfenbeinkästchen in der Online-Kollektion des Louvre<sup>10</sup> virtuell zugänglich. Außerdem findet man die Digitalisate der Illustrationen deutschsprachiger mittelalterlicher Handschriften mit Hilfe des ‚Katalogs der deutschsprachigen illustrierten Handschriften des Mittelalters‘, in dem Informationen zu den sechs Bilderhandschriften des ‚Parzivals‘ geboten werden.<sup>11</sup> Daraus lässt sich leicht eine fächerübergreifende Forschungsfrage zum Beispiel dazu entwickeln, wie unterschiedlich die Darstellung des Protagonisten als Narr ausfällt.

Während die oben angeführten Datenbanken, die in diesem hypothetischen Recherche-Entwurf neue Erkenntnisse zum ‚Parzival‘ liefern könnten, tatsächlich existieren, sind die Metadaten dieser Datenbanken bisher nicht so weit aufeinander abgestimmt, dass eine solche Form der interdisziplinär vernetzten

---

8 ‚EBIDAT – die Burgendatenbank‘, <https://www.ebidat.de> (Zugriff: 13.03.2025).

9 ‚Gothic Ivories‘, <http://gothicivories.courtauld.ac.uk> (Zugriff: 13.03.2025).

10 ‚Collections Louvre‘, <https://collections.louvre.fr/ark:/53355/cl010104045> (Zugriff: 13.03.2025).

11 ‚Deutschsprachige illustrierte Handschriften des Mittelalters‘, <https://kdih.badw.de/datenbank/start> (Zugriff: 13.03.2025).

Literaturrecherche schon realistisch ist. Trotzdem zeigt dieses Beispiel, was möglich sein könnte, wenn Datenbanken zukünftig weltweit miteinander verlinkt werden.<sup>12</sup>

Wenngleich eine so ausgefeilte Literaturrecherche mit Hilfe strukturierter Metadaten mit den heutigen Mitteln noch nicht erreichbar ist, gibt es Fortschritte im Bereich LOD: Die von der Europäischen Union geförderte Datenbank ‚Europeana‘ bezieht seit 2008 Daten aus zahlreichen europäischen Kulturinstitutionen. Für das Suchwort ‚Parzival‘ gibt ‚Europeana‘ eine beachtliche Menge an Treffern von unterschiedlichen Quellen aus: Digitalisate der Heidelberger Universitätsbibliothek, der Österreichischen Nationalbibliothek, der virtuellen Schweizer Handschriftenbibliothek ‚e-codices‘ und der slowenischen Nationalbibliothek, ältere Editionen des Wolfram’schen Romans aus der Bayerischen Staatsbibliothek, monographische Studien aus der Bodleian Library und der Deutschen Nationalbibliothek, etliche moderne bildliche Darstellungen des Stoffs sowie Theaterplakate und Szenenbilder vergangener ‚Parzival‘-Aufführungen aus Fototheken und Museen.<sup>13</sup> Damit versammelt ‚Europeana‘ Forschungsdaten, die Überlieferung, Erforschung, Illustration, Rezeption und Adaptation des ‚Parzival‘-Stoffs umfassen, und zeigt, wie weit ein Blick über den ursprünglichen Suchkontext hinaus möglich ist, wenn die frei verfügbaren Forschungsdaten kultureller Institutionen über Landesgrenzen und Disziplinen hinweg auf gemeinsamen Metadaten basieren.

Eine bibliographische Datenbank verzeichnet zunächst nur (Forschungs-)Literatur. Doch dank LOD kann sie über die fachlich naheliegende textbasierte Vernetzung mit Handschriftenportalen, Editionstexten oder Repositorien hinaus gedacht werden. Künftig werden Bibliographien mehr als eine digitale Version eines analogen Zettelkatalogs sein. Datenbanken haben das Potential, zu einer nicht-linearen, dreidimensionalen Instanz zu werden, die über ihre Verbindung mit multimodalen Daten neue Zufallsfunde, Forschungsfragen und Erkenntnisse zu generieren vermögen. Die von ‚Europeana‘ ausgegebenen Ergebnisse zeigen aber auch, dass die Zusammenführung aller möglichen Ressourcen schnell unübersichtlich wird, wenn keine nachvollziehbare (Aus-)Sortierung erfolgt: Vernetzte Zusammenhänge, die nicht qualifiziert betrachtet und strukturiert ausgegeben werden, erscheinen dann beliebig. Ein unqualifiziertes Mehr an Daten erzeugt nicht die gewünschte erkenntnisfördernde Serendipität, sondern Überforderung.

Mit Blick auf einen Teilaspekt dieser Datenflut, nämlich bibliographische Forschungsdaten, versteht sich das von der Deutschen Forschungsgemeinschaft

---

12 Das Bedürfnis nach interdisziplinärer Datenbanknutzung ist in der Mediävistik ausgeprägt. Es wird gerade deshalb als selbstverständlich erachtet, weil es zahlreiche etablierte digitale Ressourcen gibt, wie gezeigt wird von Burrichter, Gebert, Mackert u. Viehhauser 2021, S. 105.

13 ‚Europeana‘: Parzival, <https://www.europeana.eu/de/search?page=2&view=grid&query=parzival> (Zugriff: 13.03.2025).

geförderte Projekt ‚Artus digital‘<sup>14</sup> als Pilotprojekt, das einen Anfang machen will, um dieser Herausforderung zu begegnen. ‚Artus digital‘ schafft mittels der Überführung analoger bibliographischer Daten zur Artusforschung in den digitalen Raum Voraussetzungen für eine qualifizierte Vernetzung mittels LOD.

## 2 ‚Artus digital‘ als Pilotprojekt

Ziel von ‚Artus digital‘ ist es, eine als Open-Access-Angebot verfügbare Online-Bibliographie für die internationale Artusforschung zu erarbeiten, die auf der gedruckten Artusbibliographie beruht. Im Zuge des Projekts sollen außerdem Verknüpfungen zu Online-Ressourcen hergestellt und testweise Digitalisate für das Repositorium ‚The Stacks‘ des Fachinformationsdienstes (FID) Anglo-American Culture<sup>15</sup> an der SUB Göttingen erstellt werden.

Was ist zu bedenken, wenn bibliographische Daten als LOD benutzt werden sollen? Um diese Frage zu beantworten, muss die fachwissenschaftliche, bibliothekarische und technische Umsetzung gleichzeitig bedacht werden. Bisher gibt es keine standardisierten Abläufe für die Digitalisierung von Fachbibliographien. Doch um bibliographische Daten aufzubereiten, müssen sie normiert werden. Die alltägliche Erfahrung einer Hausarbeitskorrektur zeigt, wie schwierig die Herstellung von Einheitlichkeit ist, obwohl die Angaben etwa von Monographien und Zeitschriftenbeiträgen einem einheitlichen Zitierstil folgen und damit hochgradig normiert zu sein scheinen. Eine weitere Herausforderung besteht darin, dass die meisten Daten bislang nur gedruckt vorliegen oder bestenfalls als Bilddateien von Scans. Für die im akademischen Kontext geübte menschliche Wahrnehmung ist die Struktur eines bibliographischen Eintrags auch in analoger Form leicht erfassbar, wie an Abbildung 1 zu erkennen ist.

Damit die analogen Daten digitalisiert werden können, müssen sie in eine Form gebracht werden, in der sie maschinell weiterverarbeitet werden können. Beispielsweise ergeben sich aus dem Beispiel in Abbildung 1 folgende Probleme: Wo fängt der einzelne Eintrag in einer gedruckten Bibliographie an, wo hört er auf? Steht nur versehentlich ein Punkt hinter „John“ oder handelt es sich um eine Abkürzung? Gibt es einen Unterschied zwischen den Zeichen « » und „“?

---

14 Projektpartnerinnen sind die Niedersächsische Staats- und Universitätsbibliothek (SUB) Göttingen, die Verbundzentrale des Gemeinsamen Verbundkatalogs Göttingen (VZG) und die Philipps-Universität Marburg, vgl. Niedersächsische Staats- und Universitätsbibliothek Göttingen: Artus digital – Projektdetails, <https://www.sub.uni-goettingen.de/projekt-forschung/projektdetails/projekt/artus-digital/> (Zugriff: 13.03.2025).

15 The Stacks. Library of Anglo-American Culture & History, <https://thestacks.libaac.de/home> (Zugriff: 13.03.2025)

4 PARRY, JOHN. J, « *A Bibliography of Critical Arthuriana for the Year 1947* ». Offprint. (from *MLQ*, IX. 1948. 224-237.)

[Professor Parry's comprehensive annual bibliography of Arthuriana is published regularly in the June issue of *Modern Language Quarterly* and has been since the inception of that journal in 1940. The bibliography appearing in volume I for 1940, however, was prepared with the collaboration of Professor Margaret Schlauch and covers the period 1936-1939. Thus the annual bibliographies in the *Modern Language Quarterly* make a continuous series with the two volumes published under the auspices of the Modern Language Association of America : volume I by Professor Parry, *Arthurian Bibliography, 1922-1929* (1931) ; and volume II by Professors Parry and Schlauch, *Arthurian Bibliography, 1930-1935* (1936).]

**Abb. 1** | Bulletin Bibliographique de la Société Internationale Arthurienne 1 (1949). Nachdruck Amsterdam 1972, S. 13.

Um die ca. 40.000 bibliographischen Datensätze der gedruckten Bibliographie maschinenlesbar zu machen, werden sie mittels *Optical Character Recognition* (OCR) retrodigitalisiert. Das im Projekt verwendete OCR-Modell Calamari<sup>16</sup> basiert auf *Deep Learning* und wurde an einem großen Datensatz von Textbildern vortrainiert, so dass es komplexe Schriften, verschiedene Schriftarten und subtile Zeichenvariationen erkennen kann. In einer Vorverarbeitungsphase zur Verbesserung der Texterkennungsgenauigkeit werden die Bilddateien zur Kontrasterhöhung binarisiert, beschnitten und zur Beseitigung visueller Unreinheiten entrauscht sowie entzerrt. Im Anschluss erfolgen zwei Schritte: Zunächst muss die Segmentierung (das Seitenlayout) richtig erkannt werden, erst dann werden die einzelnen Zeichen erfasst. Sowohl Segmentierung als auch Texterkennung müssen in Stichproben überprüft und gegebenenfalls manuell nachkorrigiert werden. Bei der Segmentierung wird beispielsweise häufiger nicht jeder einzelne bibliographische Eintrag auch als solcher erkannt. In Abbildung 2 wurde Eintrag Nr. 15 fälschlicherweise als zwei separate Segmente interpretiert. Die maschinelle

<sup>16</sup> Vgl. GitHub: Calamari-OCR, [https://github.com/Calamari-OCR/calamari\\_models/tree/master/uw3-modern-english](https://github.com/Calamari-OCR/calamari_models/tree/master/uw3-modern-english) (Zugriff: 13.03.2025). Das Modell nutzt die im OCR-D-Projekt (<https://ocr-d.de/de/about>, Zugriff: 13.03.2025) entwickelten OCR-Prozessschritte nach. Dazu gäbe es zahlreiche Alternativen; empfehlenswert für die Orientierung ist der OCR-Recommender: <https://wiki.bib.uni-mannheim.de/limesurvey/index.php/996387?lang=de> (Zugriff: 13.03.2025).

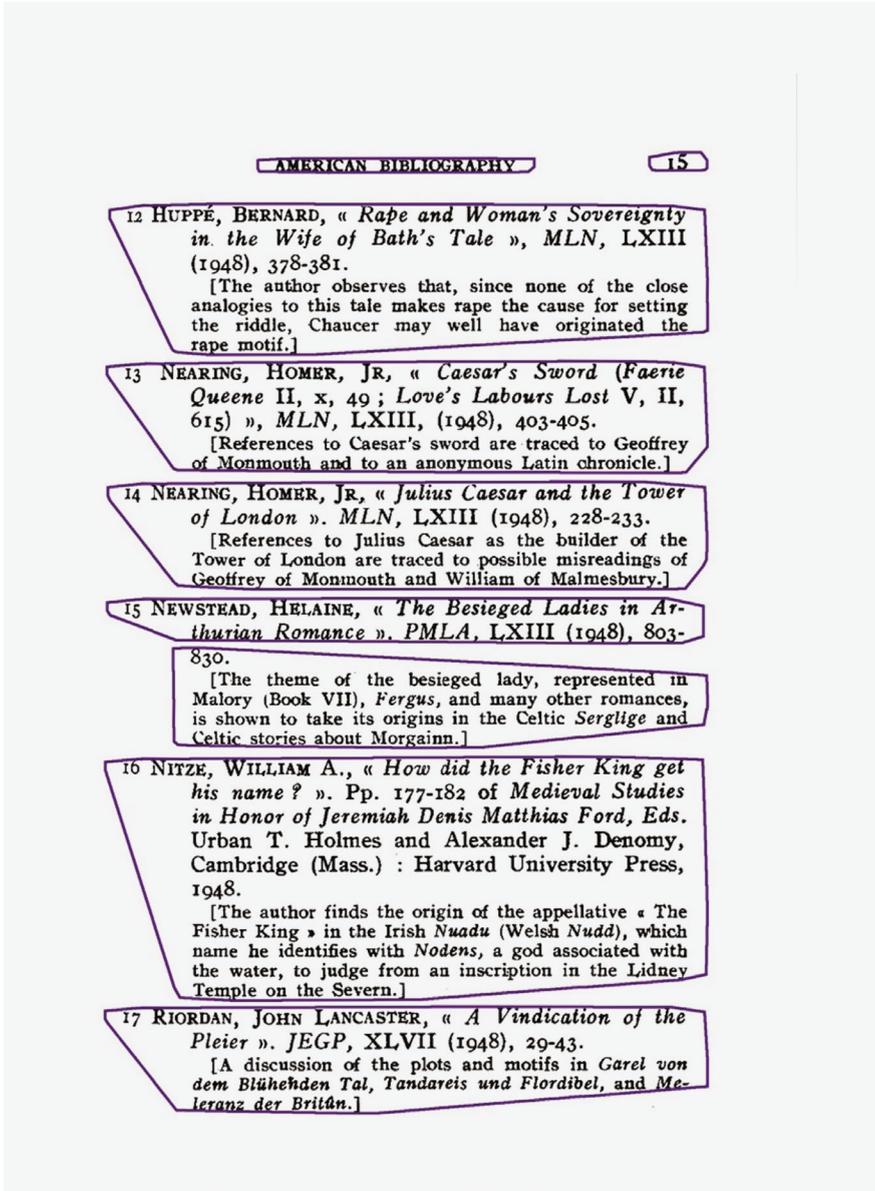


Abb. 2 | Bulletin Bibliographique de la Société Internationale Arthurienne 1 (1949). Nachdruck Amsterdam 1972, S. 13.

Segmentierung wird deshalb im Projekt vor der Weiterverarbeitung vom fachkundigen Projektteam in der Open-Source-Software OCR4all korrigiert.<sup>17</sup>

Weil das Modell außerdem an englischen Korpora vortrainiert wurde, gestaltet sich die Texterkennung bei sprachspezifischen Sonderzeichen in der mehrsprachigen Artusbibliographie mitunter schwierig. Darum wird das OCR-Modell im Laufe des Projekts mit verschiedenen Parametern iterativ weiterentwickelt.

Sobald die Volltextdigitalisate aller gedruckten Bände vorliegen, müssen sie in einem Datenmodell abgebildet werden. Im Projekt wird ChatGPT verwendet, um bibliographische Informationen zu interpretieren und in ihre Bestandteile zu zerlegen (*Parsing*), dies verbessert die Effizienz und Genauigkeit des Datenextraktionsprozesses. Darüber hinaus wandelt das KI-Tool die extrahierten Ergebnisse in ein strukturiertes Format um, so dass die bibliographischen Daten weiterverarbeitet werden können.

Die gedruckten Bände der Artusbibliographie werden durch einen Schlagwortindex erschlossen. Dessen Einträge verbinden die einzelnen Literaturtitel durch ein inhaltliches Kriterium miteinander und machen sie – analog wie digital – gemeinsam auffindbar. Die gedruckten Verzeichnisse erleichtern zwar die analoge Suche im jeweiligen Band, sind jedoch von Band zu Band unterschiedlich, etwa weil sich über die Jahrzehnte ihrer Entstehung unterschiedliche Forschungstrends auf die Schlagwortlisten auswirkten oder auch einfach weil zwei Forscher\*innen einen ähnlichen Sachverhalt unterschiedlich bezeichneten. Aufgrund ihrer Uneinheitlichkeit sind sie für eine digitale Suche über alle Bände hinweg nicht brauchbar. Für die Online-Bibliographie müssen diese Tausende von Schlagwörtern reduziert werden, damit sie einerseits die zahlreichen Einträge aus den 68 bereits erschienenen Bänden gemeinsam erschließen und andererseits die Schlagwörter auch bei zukünftigen Einträgen verwendbar sind. Dazu wird im Laufe des Projekts ein reduziertes Vokabular von wenigen Hundert Schlagwörtern erstellt, das bei sämtlichen vergangenen und künftigen Jahrgängen zur Anwendung kommen wird. Dieses Vokabular zu erstellen, erfordert eine sorgfältige Auswertung der bibliographischen Daten. Dazu muss die jeweilige fachliche Bandbreite der Schlagwörter abgeschätzt und es müssen ähnlich formulierte Schlagwörter unter einem gemeinsamen Oberbegriff zusammengeführt werden, so dass anhand von Statistiken unikale Verwendungen und ‚Irrläufer‘ aussortiert werden können. Weil Forschungsinteressen und ihre Bezeichnung sich im Laufe der Zeit ändern, ist für die Aktualisierung der Schlagwörter ein beständiger Austausch mit der internationalen Forschungscommunity unabdingbar, etwa durch Einbezug der Expertise des bibliographischen Teams. Allzu komplex oder gar mehrstufig darf das Vokabular allerdings nicht werden, da seine Verwendung innerhalb eines Teams, das sich

---

17 OCR4all, <https://www.ocr4all.org/> (Zugriff: 13.03.2025).

The screenshot shows a search interface for the 'INTERNATIONAL ARTHURIAN SOCIETY' database. The search term 'parry' has been entered, and the results are sorted by 'Erscheinungsjahr'. The search results show one entry with the following details:

Suchgeschichte	Kurzliste	<b>Titelanzeige</b>
Ihre Aktion suchen [und] ([ALL] Alle Wörter) parry   1 Treffer		
Speichern/Drucken   Merken		
PPN:	000007234 <a href="#">Kritisch</a>	
Aufsatztitel:	Arthur and Possibility : The Philosophy of Lagamon's Arthuriad	
Verfasser:	Parry, Joseph [VerfasserIn]	
Einheitsachtitel:	Layamon's Brut#bibliographer: Ann Howey	
Sprachen:	Englisch	
Anmerkung:	"Critical and Historical Studies" [Forschungsliteratur]	
Bibliogr. Zusammenhang:	Enthalten in: <i>Arthuriana</i> . - Dallas, Texas : Southern Methodist Univ. ; Scriptorium Press West Lafayette Bd. 26 (2016), 1, S. 60-75	
Schlagwörter:	*history	
Inhalt:	In the <i>Brut</i> 's Arthuriad, Lagamon becomes something of a philosopher of history. In response to Wace's tragic Arthur, Lagamon's becomes an expression of possibility. History becomes a way of understanding the structure of human experience not only as it has been lived, but also as it may be lived. (JP)	
Identifizier:	DOI: 10.1353/art.2016.0004	
Link zum Digitalisat:	<a href="https://doi.org/10.1353/art.2016.0004">https://doi.org/10.1353/art.2016.0004</a>	

**Abb. 3** | Beispielseintrag aus der Artus-Testdatenbank. <https://kxp.k10plus.de/DB=1.85/PPNSET?PPN=000007234> (Zugriff: 13.03.2025).

nicht hauptberuflich mit der Bibliographie beschäftigt, unrealistisch ist. Vielmehr ist das Ziel, ein einfach verwendbares, nicht erklärungsbedürftiges Vokabular für die künftige Arbeit der internationalen Bibliograph\*innen zu erstellen.

Nach dem OCR-Prozess, der Datenextraktion und der Erstellung der Schlagwortlisten für Autoren, Werke und Sachen sind die Daten so weit aufbereitet, dass sie in einer bibliographischen Rechercheumgebung angezeigt werden können, über die Nutzende auf sie zugreifen, sie durchsuchen, filtern und exportieren können. Es erweist sich als großer Vorteil, dass mit der Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG) eine Projektpartnerin gewonnen werden konnte, die bereits OPAC-Systeme verwendet, die im bibliothekarischen Kontext gängig sind und für die bereits vorbildliche Umgebungen<sup>18</sup> vorliegen, die zum Beispiel für die Online-Präsenz der ‚Leibniz-Bibliographie‘<sup>19</sup> und der ‚Bibliographie des Musikschritfttums online‘<sup>20</sup> eingesetzt werden. Die Such- und Filteroptionen werden an die Projektanforderungen angepasst, die in engem Austausch mit Mitgliedern der Artusgesellschaft ermittelt werden. Abbildung 3 zeigt beispielhaft einen beliebigen bibliographischen Eintrag in der Artus-Testdatenbank. Die einzelnen Datenfelder

<sup>18</sup> Diese basieren auf der Software *Central Bibliographic System* von OCLC (Online Computer Library Center), <https://www.oclc.org/de/home.html> (Zugriff: 13.03.2025) und werden von diesem kontinuierlich weiterentwickelt.

<sup>19</sup> ‚Leibniz-Bibliographie‘, <https://www.leibniz-bibliographie.de/> (Zugriff: 13.03.2025).

<sup>20</sup> ‚Bibliographie des Musikschritfttums online‘, <https://www.musikbibliographie.de/> (Zugriff: 13.03.2025).

sind nun normiert und die Forschungsdaten können daher nicht nur von Menschen genutzt, sondern auch von Maschinen weiterverarbeitet werden.

Dank des Fachwissens der Community kann der Datenbestand fortwährend um neue Daten erweitert werden. Von der Artusgesellschaft bestellte bibliographische Sekretär\*innen erfassen in einer ersten Testphase neue Schriften der Jahre 2023 und 2024 mit dem Literaturverwaltungsprogramm Zotero.<sup>21</sup> Dieses Programm ist weit verbreitet und bietet sich für solche Projekte deshalb an, weil es Open Source angeboten wird, auch von Wissenschaftler\*innen ohne bibliothekarische Vorkenntnisse benutzt werden kann und darin vielfache Vernetzungen vorgesehen sind. Allerdings ist Zotero für LOD unbrauchbar, denn die Daten werden lediglich in Freitextfelder eingetragen, nicht aber weiterverarbeitet. Deshalb setzt ‚Artus digital‘ neben Zotero auch den OPAC ein: Die Daten in Zotero werden nicht öffentlich angezeigt, sondern nach der Eingabe über eine eigene Import-Schnittstelle in den OPAC eingespielt und an die Erfordernisse des Datenmodells angepasst. Anders als Zotero bietet die im Laufe des Projekts entwickelte bibliographische Datenbank die Möglichkeit, die eingescannten Daten zu LOD aufzuwerten, so dass damit weitere Weichen für die oben beschriebene Vision der vernetzten wissenschaftlichen Recherche der Zukunft gestellt werden können.

### 3 Hinterm Bücherregal

Welcher Aufwand im Projekt ‚Artus digital‘ nötig ist, um eine gedruckte Bibliographie in eine durchsuchbare Datenbank zu überführen, kann am Beispiel von Abbildung 1 gesehen werden. Im Druck war es unproblematisch, einen Sonderdruck aus einer Zeitschrift anzugeben; in einer Datenbank wäre die Normierung dieser Angabe nur mit einigem Aufwand möglich, da die Angabe vom üblichen Datenmodell abweicht. Zur Vereinfachung reicht es aus, die Zeitschrift anzugeben, in welcher der Aufsatz erschienen ist, und die Angabe in ein einheitliches Stylesheet zu überführen. Unter Verwendung der Zitiergepflogenheiten von ‚Das Mittelalter‘ sähe die Angabe beispielsweise so aus:

**Parry, John J.:** A Bibliography of Critical Arthuriana for the Year 1947. In: *Modern Language Quarterly* 9 (1948), S. 224–237.

Mit der Ausgabe dieses Titels hätte eine bibliographische Datenbank an erster Stelle ihre Aufgabe erfüllt. Als Arbeitsinstrument der Wissenschaftskommunikation

---

21 Der hier beschriebene Workflow spiegelt den Arbeitsstand Ende 2024. Im Projekt werden seitdem neue Optionen diskutiert, die einen Wechsel der gewählten Datenbanksysteme denkbar machen.

würde diese Angabe zu einem Objekt führen, nämlich zu einem Beitrag in einer bestimmten Zeitschriftennummer, die beispielsweise in den Magazinen der Universitätsbibliothek Göttingen unter der Signatur ‚8 Z LIT 417‘ aufbewahrt wird. Die Daten würden also in erster Linie auf ein physisches Objekt zeigen.

Bei einer bibliographischen Datenbank besteht jedoch das Interesse, die verzeichneten Objekte direkt aus ihr heraus aufrufen zu können. Als erster Schritt wären diese Daten folglich um Links auf Digitalisate des Beitrags zu erweitern, wie eines davon von der Duke University Press online hinter einer Bezahlschranke für Abonnent\*innen angeboten wird.<sup>22</sup> Eine Verlinkung mit Repositorien, die diese Objekte anbieten, ist fast schon Standard. Dadurch wäre gegenüber den gedruckten Bänden viel gewonnen und ein nützliches Arbeitsinstrument geschaffen. Nur um LOD handelt es sich bei diesen bibliographischen Daten noch nicht und außerdem bleibt eine solche Datenbank noch immer hinter ihren Möglichkeiten zurück. Die Titelangaben werden bloß als Metadaten behandelt, die auf materielle Buch-Objekte verweisen. Dabei stellen sie ihrerseits Forschungsdaten dar, die weiterverarbeitet und erforscht werden können. Was ist also zu bedenken, damit aus normierten bibliographischen Daten tatsächlich LOD werden?

Die Zusammenstellung einer thematischen Bibliographie war immer schon eine intellektuelle Leistung, aber um als LOD zu fungieren, genügt es nicht, sie in eine Datenbank zu stellen. Eine der zahlreichen Hürden bei der Digitalisierung gedruckter Bibliographien ist die Überführung des eingelesenen Texts in ein offenes Datenmodell. Als Hilfestellung für die Modellierung solcher Daten wurden 2016 die ‚FAIR Guiding Principles for Scientific Data Management and Stewardship‘ vorgeschlagen.<sup>23</sup> Das Akronym FAIR steht für die Anforderungen, die an die Daten gestellt werden: Sie sollen *Findable*, *Accessible*, *Interoperable* und *Reusable* sein. Die FAIR-Prinzipien geben auch für den Umgang mit bibliographischen Daten eine angemessene Richtschnur vor.

Schon das Auffinden der Daten bereitet Probleme, denn was angesichts von gängigen Suchmaschinen nach einer alltäglichen Aufgabe klingt, hat seine Tücken im Detail. Zunächst muss die Bibliographie selbst gefunden werden. Bei Suchmaschinen wie Google, Bing oder Baidu liegt eine Schwierigkeit schon in den Besitzverhältnissen. Die Portale befinden sich in privaten Händen und haben mit den Seiten, die sie verzeichnen, nichts zu tun. Wie aber kann garantiert werden, dass eine heute aufgebaute Datenbank auch in 50 Jahren noch auffindbar ist? Man wird sich zumindest auch im geisteswissenschaftlichen Umfeld frühzeitig um die Auffindbarkeit durch verschiedenartige Suchverfahren bemühen müssen. Ferner ist ein Verfahren zu entwickeln, welches zur passenden Suchmaschine

---

22 Vgl. <https://doi.org/10.1215/00267929-9-2-224> (Zugriff: 13.03.2025).

23 Vgl. Petz 2023, S. 215.

leitet, denn bereits heute gibt es weit mehr Fachressourcen im Netz, als selbst die Expert\*innen-Community nutzt.

Auffindbar sein muss aber nicht nur die Datenbank, sondern auch das Objekt, ob nun in physischer oder digitaler Form. Dass ein Buch auch über einen längeren Zeitraum im Bestand einer Bibliothek bleibt und dort zugänglich ist, setzt einen Sammlungsantrag der Bibliothek voraus; diese wird auch dauerhaft Instrumente zum Auffinden der Medien bereitstellen wollen. Damit ein Artikel in digitaler Form gefunden werden kann, ist nicht nur eine Suchmaschine wichtig, sondern auch eine eindeutige Adressierung des Objekts, wie sie zum Beispiel oben in Form eines DOI (*Digital Object Identifier*) vergeben wurde. Indies: Ob die aktuell als persistent geltenden Adressen tatsächlich von Dauer sind, wird sich erst langfristig zeigen.

Mit den DOI wurde ein wichtiger Schritt in die Richtung einer eindeutigen Identifizierung von Objekten (in diesem Fall: elektronischer Publikationen) gegangen, denn für die Auffindbarkeit ist auch entscheidend, dass allen Objekten weltweit nur eine einzige, eindeutige ID zugewiesen wird. Erstaunlicherweise gibt es im bibliothekarischen Bereich bislang keine globalen Standards: Die allseits bekannte ISBN (*International Standard Book Number*) ist für diesen Zweck nutzlos, denn sie dient in erster Linie den Anforderungen im Verkauf und wird darüber hinaus nur für selbstständig erschienene Publikationen vergeben, so dass ältere Studien, hochschulintern veröffentlichte Schriften, Zeitschriftenaufsätze und Beiträge in Sammelbänden nicht mit ihrer Hilfe gefunden werden können. Unveränderte Nachdrucke erhalten ebenfalls nicht immer eine neue ISBN. Der inzwischen verbreitete DOI ist dem digitalen Bereich vorbehalten. Nummern von Verbundkatalogen (z. B. PPN oder OCLC) werden zwar bibliothekarisch genutzt, sind in der Breite aber unbekannt.

Noch in einem dritten Sinne ist das Auffinden schwierig: Der Autor des oben genannten Beispiels lautet „Parry, John. J.“. Der gebürtige Brite Parry war Professor an der Universität von Illinois und machte sich verdient um die Artusforschung, weil er schon vor Gründung der Internationalen Artusgesellschaft Bibliographien und Editionen vorlegte. Man sollte meinen, er wäre in einem Bibliothekskatalog leicht zu finden. Doch gibt man seinen Namen beispielsweise im ‚Gemeinsamen Verbundkatalog‘ (GBV) ein, findet man zahlreiche Menschen mit dem Namen John Parry. Auch ein Experte oder eine Expertin wird länger suchen müssen, um den richtigen Menschen zu identifizieren, dessen Vorname(n) als John, John J. oder John Jay verzeichnet sind. All diese Varianten machen eine maschinelle Verarbeitung schwierig. Wie also können die Daten so strukturiert werden, dass ausschließlich Suchergebnisse von Publikationen einer einzigen Person ausgegeben werden? Die Antwort liegt in der Verwendung von Normdaten.<sup>24</sup> Diese stellen eindeutige IDs unter anderem für Personen bereit. Die in Deutschland am weitesten verbreitete

---

24 Ausführlicher und einführend: Busch u. Müller 2023.

Normdatei ist die ‚Gemeinsame Normdatei‘ (GND), die von der Deutschen Nationalbibliothek zur Verfügung gestellt wird. In ihr ist John Parry die ID 128596635 zugewiesen. Mit dieser Nummer ist man in der Lage, ihn in allen Namensvarianten zu finden und ihn von anderen Menschen gleichen Namens zu unterscheiden. Wenn nun also in einer Datenbank nicht nur der Name, sondern auch die GND-ID hinterlegt ist, dann ist es möglich, sämtliche Publikationen Parrys gemeinsam anzuzeigen, ganz egal, wie genau der Name geschrieben wird – und zugleich alle anderen Personen gleichen Namens auszuschließen. Solche Normdateien haben sich inzwischen in den *Digital Humanities* weitgehend durchgesetzt. Außerdem besitzen viele Wissenschaftler\*innen eine ORCID (*Open Researcher and Contributor ID*), die sie mit ihren Publikationen verknüpfen. Da sich die Wissenschaftler bei ORCID selbst registrieren müssen und es diese Identifikationsmöglichkeit erst seit 2010 gibt, verfügen fast nur heute noch lebende Forscher\*innen über eine ORCID, zudem ist diese Registrierung freiwillig und wird daher nicht von allen genutzt.

Eng mit der Auffindbarkeit verbunden ist das zweite der FAIR-Prinzipien, die Zugänglichkeit der Daten, denn noch schwerer vorhersehbar als die zukünftigen Suchmaschinen ist die Art und Weise der Datenpräsentation selbst. Die Daten müssen an einem Ort abgelegt werden, der dauerhaft und direkt erreichbar ist – also ohne dass etwa eine Bezahlschranke zu überwinden ist, wie es bei der in der Duke University Press erscheinenden Zeitschrift ‚*Modern Language Quarterly*‘ der Fall ist – und der auch eine Schnittstelle anbietet, über die die Daten dauerhaft zu beziehen sind. Im schlimmsten Fall wird eine einmal für geeignet empfundene Plattform nach wenigen Jahren deaktiviert und die Daten sind nicht mehr verfügbar. Die FAIR-Prinzipien sehen für diesen Fall vor, dass wenigstens Metadaten an anderem Ort erhalten bleiben sollen, die die verlorenen Daten beschreiben. Indes ist es ein rekursives Problem, inwiefern Metadaten einfacher aufzubewahren sein sollen als die Daten selbst. Zumindest wird man recht daran tun, sich an Datenstandards zu halten, die sich im aktuellen Moment durchgesetzt haben, und nicht Formate zu entwerfen, die nur für den Projektkontext gelten.

Während die ersten beiden FAIR-Prinzipien Anforderungen an den zeitlichen und formalen Zugang formulieren, wird mit dem dritten und vierten Prinzip, *interoperable* und *reusable*, das technische und rechtliche Zusammenwirken mit anderen Ressourcen fokussiert.<sup>25</sup> Damit Daten zwischen den Datenbanken automatisiert ausgetauscht werden können, ist erstens eine Schnittstelle erforderlich, zweitens müssen die Daten kompatibel strukturiert sein und schließlich müssen sie diese Struktur in Metadaten preisgeben. Darüber hinaus sind für die Wiederverwendbarkeit freie Lizenzen erforderlich, die eine Verwendung über den primären Datengeber hinaus ermöglichen. Die öffentlichen Förderinstitutionen

---

25 Dazu etwa Ben Aouicha, Turki u. Hadj Taieb 2023, S. 2.

fordern inzwischen weitgehend, dass die von ihnen geförderten wissenschaftlichen Publikationen im Open Access und damit für alle online frei zugänglich erscheinen.

Die Erfüllung der vier Forderungen von FAIR steht und fällt mit der Gestaltung des Datenmodells. Im bibliographischen Bereich sind die Datenmodelle weitgehend standardisiert und brauchen nicht neu entwickelt zu werden.<sup>26</sup> Beispielsweise ist es im akademischen Bereich üblich, bei der bibliographischen Beschreibung von Zeitschriftenaufsätzen keine Zeitschriftenherausgeber\*innen anzugeben. Auch Verlagsorte werden bei Zeitschriften nicht genannt. Dort wo die Nennung des Ortes üblich ist, etwa in der Germanistik bei Monographien, werden sie in einer Datenbank stets im selben Tabellenfeld abgelegt, für eine automatisierte Weiterverwendung normiert (zum Beispiel durch einheitliche Interpunktion zwischen zwei Ortsnamen) und von anderen Orten gleichen Namens eindeutig unterschieden (etwa durch Normdaten). Dieser Aufwand zahlt sich nicht nur in der einfachen Durchsuchbarkeit aus, sondern vor allem in der Weiterverarbeitung. Die Daten des Projekts stehen frei zur Verfügung. Sie können problemlos in weitere Literaturverwaltungsprogramme (Citavi, Zotero etc.) übertragen werden, so dass sie von allen, die sich für Artusforschung interessieren, weiter nutzbar sind. Darüber hinaus ist die Verwendung der Daten aus der Artusbibliographie als eigenständiges Forschungsobjekt möglich. Insbesondere für wissenschaftsgeschichtliche Fragestellungen sind die Titeldaten aufschlussreich und bislang nie ausgewertet worden. Als Beispiel sei noch einmal die eingangs genannte Frage nach den Ortsnamen in Wolframs ‚Parzival‘ aufgegriffen. Wenn man die bibliographischen Daten aus 70 Jahren überblickt, wird ersichtlich, in welcher Zeit und unter welchen Gesichtspunkten Ortsnamen von Interesse waren, so dass sich darin zugleich eine Entwicklung der Artusforschung erkennen lässt. Damit ließen die Daten sich auch daraufhin auswerten, in welchen Primärtexten bislang noch keine Ortsnamen untersucht wurden. Auf diese Weise als Forschungsdaten befragt, führen die bibliographischen Daten nicht nur zu weiterer Literatur, sondern geben ihrerseits Auskunft darüber, was die Forschung im Blick hatte – und was sie übersehen hat.

Die bibliographischen Datenmodelle sind allerdings gerade im bibliothekarischen Bereich derart kompliziert, dass eine langwierige Ausbildung erforderlich ist, damit man sie angemessen bedienen kann. Gerade für LOD ist deshalb die Hürde enorm hoch, weil die Austauschbarkeit ein hohes Maß an Arbeit erfordert, das nicht in jedem Fall verlangt werden kann. Die im Wissenschaftsbetrieb verbreiteten Literaturverwaltungsprogramme sind nicht auf die erforderliche Normierung oder gar die Verwendung von Normdaten ausgelegt.<sup>27</sup>

Aus dem Projekt ‚Artus digital‘ ergeben sich aber auch Erkenntnisse über den bibliographischen Bereich hinaus. Die Beschäftigung mit Bibliographien ist so weit

---

<sup>26</sup> Vgl. z.B. Danowski u. Pohl 2013.

<sup>27</sup> Vgl. Zumstein u. Stöhr 2015.

institutionalisiert, dass auf zahlreiche Vorarbeiten zurückgegriffen werden kann. Das Datenmodell musste nicht neu entwickelt, sondern nur angepasst werden. Vor allem ist es im bibliothekarischen Kontext umfassend eingeführt, so dass es nicht bekanntgemacht oder durchgesetzt werden muss. Entfernt man sich vom bibliographischen Bereich, dann ist das nicht der Regelfall. Beispielsweise gibt es kein Standardformat für die Erfassung mittelalterlicher Handschriften: Jedes kodikologische Projekt ist darauf angewiesen, ein Modell zu entwickeln, was den Austausch erschwert.<sup>28</sup> Auch für die Verwendung der Software kann bei ‚Artus digital‘ auf einen hohen Institutionalierungsgrad zurückgegriffen werden, bei dem Standards etwa für Programme und Datenmodell bereits entwickelt sind, denn jede eigenständig in einem Projekt entwickelte Software wird kaum langfristig, das heißt über die Projektlaufzeit hinaus, zugänglich sein. Es ist hingegen anzunehmen, dass Bibliotheken auch noch in fünfzig Jahren mit Nachfolgesystemen der jetzigen Datenbank arbeiten werden und eine Übertragung möglich sein wird. Gleiches wäre zum Beispiel für digitale Editionen wünschenswert, bei denen sich bislang kein standardisierter Weg der Präsentation abzeichnet.

‚Artus digital‘ mag heute als Projekt wie ein Sonderfall erscheinen, der wenig mit den Unwägbarkeiten des akademischen Alltags gemein hat; doch dem bewilligten Antrag ging eine fast zehnjährige Vorbereitungszeit voran, in der sich einige eingeschlagene Wege als Holzwege herausstellten. Statt einer Zusammenfassung sei daher das Konzept in fünf Empfehlungen formuliert:

1. Anbindung an bestehende Communities. Es ist eine Selbstverständlichkeit der wissenschaftlichen (Projekt-)Arbeit, dass man sich Gleichgesinnte sucht. Die Internationale Artusgesellschaft sammelt seit jeher Fachliteratur und wird das auch künftig tun; damit ist sie aber als Forschungscommunity kein Sonderfall. Ein wesentlicher Vorteil einer solchen Community-Anbindung besteht in der langfristigen Verfügbarkeit der Daten, für die sich dann auch andere interessieren. Darum sollte eine Anbindung an bestehende Communities sowie an deren Programme und Standards angestrebt werden.
2. Verwendung vorhandener Programme. Es ist einigermaßen einfach und erscheint auch überaus attraktiv, für jedes Projekt ein eigenständiges Programm oder eine eigene Datenbank zu entwickeln. Doch selbst wenn

---

<sup>28</sup> Zahlreiche Initiativen arbeiten an einem Standard. Zu erwähnen wäre etwa das Datenmodell des ‚Handschriftenportals‘ oder die *Text Encoding Initiative* (TEI), die ein Element `<MsDesc>` entwickelt hat. Außerdem arbeitet eine AG Handschriften an der Arbeitsstelle für Standardisierung (AfS) der Deutschen Nationalbibliothek seit einigen Jahren an einer Integration von Handschriftenbeschreibungen in den Katalogisierungsstandard RDA (*Resource Description and Access*). Die Initiative für die Erstellung eines *International Standard Manuscript Identifier* (ISMI) ist leider ins Stocken geraten. Bislang hat sich keine Initiative durchsetzen können.

man über ausreichend Projektmittel verfügt, um ein eigenes Programm zu entwickeln, und sogar noch das Glück hat, gutes Personal gefunden zu haben, das diese Programme schreiben kann, so ist dennoch das Risiko groß, dass die Programme schon kurz nach Projektende nicht mehr funktionieren. Man sollte auf Programme zurückgreifen, die unabhängig vom Projekt entwickelt wurden und bei denen auch von anderer Seite Interesse an einer Weiterentwicklung besteht. Die Entscheidung, bei ‚Artus digital‘ Zotero und den OPAC zu nutzen, ist das Ergebnis eines lange ausgehandelten Kompromisses. Dieser wurde auch deshalb akzeptiert, weil das Projektteam bereits Erfahrungen mit einer Vorläufer-Datenbank gemacht hatte, die eigens für die Artusbibliographie erstellt, aber nicht instand gehalten werden konnte.

3. Verwendung vorhandener Datenstandards. Analog zur Wahl des Programms möchte man in einem innovativen Projekt sicherlich auch neue Datentypen sammeln. Dies geht jedoch mit der Gefahr einher, dass die Modelle bald nicht mehr überschaubar sind. Auch hier ist zu überlegen, ob sich das eigene Forschungsinteresse nicht doch über eine Anpassung an einen vorhandenen Standard verfolgen lässt.
4. Verwendung von Normdaten. Erst durch die eindeutige Identifikation von Objekten können die dazu gehörigen Metadaten austauschbar gemacht werden. Dazu hilft es, wenn Projekte frühzeitig die Verwendung von Normdaten vorsehen.
5. Zuletzt: *User Experience*. Viele Konzepte im Zusammenhang mit LOD sind unbekannt und werden es auch bleiben. Wer immer Portale oder Datenbanken baut, muss auch die intendierten Nutzenden im Blick haben, die die Konzepte nicht kennen. Man muss nicht im Einzelnen wissen, was Normdaten sind, um sie etwa im Rahmen einer Literaturrecherche zu verwenden.

LOD zeigt sich als ein Konzept, das auf vielfältige Weise umgesetzt werden kann und das Kreativität erfordert. Wie eine Edition vor der Herausforderung steht, aus einer Vielzahl an Möglichkeiten eine angemessene Repräsentation eines Textes zu wählen, und es auch unzählige Möglichkeiten gibt, ein Bibliotheksregal zu sortieren, so sind auch Daten nicht bloß auf eine einzige Art und Weise modellierbar. Entscheidungen haben Folgen für die weitere Arbeit mit Texten, Regalen und Daten. Die Ergebnisse der Nutzenden sind abhängig von der Qualität der Arbeit und von der Art der Präsentation. Die Bereitstellung von LOD ist deshalb Grundlagenarbeit.

## Literaturverzeichnis

- Ben Aouicha, Mohamed, Houcemeddine Turki u. Mohamed Ali Hadj Taieb:** Editorial: Linked Open Bibliographic Data for Real-Time Research Assessment. In: *Frontiers in Research Metrics and Analytics* 8 (2023). <https://doi.org/10.3389/frma.2023.1275731> (Zugriff: 13.03.2025).
- Borek, Luise:** Arthurische Pferde als Bedeutungsträger. Eine Fallstudie zu ihrer digitalen Klassifizierung (*Zeitschrift für deutsches Altertum und deutsche Literatur*, Beihefte 43). Stuttgart 2023.
- Bulletin Bibliographique de la Société Internationale Arthurienne* 1 (1949). Amsterdam 1972 (Nachdruck).
- Burrichter, Brigitte, Björn Gebert, Christoph Mackert u. Gabriel Viehhauser:** Digitale Mediävistik. In: Wolfram Drews, Matthias Müller u. Regina Toepfer (Hgg.): *Mediävistik 2021. Positionen, Strategien, Visionen (Das Mittelalter. Perspektiven mediävistischer Forschung 26)*. Heidelberg 2021, S. 101–117. <https://doi.org/10.17885/heiup.mial.2021.1.24312> (Zugriff: 13.03.2025).
- Busch, Nathanael u. Diana Müller:** Normdaten in den Geisteswissenschaften. *Frequently Asked Questions*. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 53 (2023), S. 781–796. <https://doi.org/10.1007/s41244-023-00295-1> (Zugriff: 13.03.2025).
- Danowski, Patrick u. Adrian Pohl (Hgg.):** (Open) Linked Data in Bibliotheken. Berlin, Boston 2013.
- Erdelez, Sandra, Jannica Heinström, Stephann Makri u. a.:** Research Perspectives on Serendipity and Information Encountering. In: *Proceedings of the Association for Information Science and Technology* 53 (2016). <https://doi.org/10.1002/pras.2016.14505301011> (Zugriff: 13.03.2025).
- Kennedy, Ian, Dean Whitehead u. Debra Ferdinand-James:** Serendipity: A Way of Stimulating Researchers' Creativity. In: *Journal of Creativity* 32 (2022). <https://doi.org/10.1016/j.yjoc.2021.100014> (Zugriff: 13.03.2025).
- Petz, Georg:** Linked Open Data. Zukunftsweisende Strategien. In: *Bibliothek Forschung und Praxis* 47 (2023), S. 213–222. <https://doi.org/10.1515/bfp-2023-0006> (Zugriff: 13.03.2025).
- Zumstein, Philipp u. Matti Stöhr:** Zur Nachnutzung von bibliographischen Katalog- und Normdaten für die persönliche Literaturverwaltung und Wissensorganisation. In: *ABI Technik* 35 (2015), S. 210–221. <https://doi.org/10.1515/abitech-2015-0037> (Zugriff: 13.03.2025).

# Automatische intratextuelle Kohärenzanalyse am Beispiel des ‚Marienlebens‘ Philipps von Seitz

## Kontakt

**Dr. Stefan Tomasek,**  
Universität Würzburg,  
Lehrstuhl für deutsche Philologie,  
ältere Abteilung, Am Hubland,  
97074 Würzburg,  
stefan.tomasek@uni-wuerzburg.de

**PD Dr. Christian Reul,**  
Universität Würzburg,  
Zentrum für Philologie und Digitalität,  
Emil-Hilb-Weg 23,  
Campus Hubland Nord,  
97074 Würzburg,  
christian.reul@uni-wuerzburg.de

**Franziska Schulte,**  
Universität Würzburg,  
Lehrstuhl für deutsche Philologie,  
ältere Abteilung, Am Hubland,  
97074 Würzburg,  
franziska.schulte@stud-mail.  
uni-wuerzburg.de

**Kiara Hart,**  
Universität Würzburg,  
Zentrum für Philologie und Digitalität,  
Emil-Hilb-Weg 23,  
Campus Hubland Nord,  
97074 Würzburg,  
kiara.hart@stud-mail.uni-wuerzburg.de

**Abstract** The article presents a process for automatically identifying and systematising linguistic repetition structures in Middle High German texts, using Philipps von Seitz ‘Marienleben’ as a case study. In a second step, the results are categorised in genre contexts. Additionally, an OCR process (OCR4all) is described which can be used to create the data basis even in the absence of a sufficient edition basis, as is currently the case with ‘Marienleben’.

**Keywords** Automatic Coherence Analysis; Comparison of Genre Contexts; Digital Medieval Studies; Marienleben; OCR

*Marien müter auch vrov anne |  
Mit hern ioachim ir manne | kegen ir  
töchter giengen | Mit grözen freuden  
fî enphiengen | Si sprachen wilkom  
kint marie | vilkom tochter magt vri |  
wilkom vrowe kuniginne | Ifrahelifh  
löferinne (v. 10197–10204).<sup>1</sup>*

- <sup>1</sup> „Auch die Mutter Mariens, die Herrin Anna, ging ihrer Tochter gemeinsam mit dem Herren Joachim, Annas Ehemann, entgegen. Sie begrüßten sie mit großer Freude. Sie sagten: ‚Willkommen Maria, mein Kind, willkommen Tochter, unverheiratete Jungfrau, willkommen Herrin und Königin, Erlöserin Israels!‘“ Den Zitaten aus dem ‚Marienleben‘ Philipps von Seitz liegt die Handschrift P, Pommersfelden, Gräfl. Schönbornsche Schlossbibl., Cod. 46 (2797) zugrunde. Vgl. zu dieser Handschrift Gärtner 1978, S. 160–164, und <https://handschriften.census.de/9432> (Zugriff: 08.11.2024). Die Versangaben entsprechen der Zählung der HTR-Transkription der

So wird die in den Himmel aufgestiegene Maria im ‚Marienleben‘ Philipps von Seitz von ihren Eltern Anna und Joachim begrüßt.<sup>2</sup> Eine Grußformel, die den Rezipient:innen des Textes nicht unvertraut ist, denn ganz ähnlich hatte die Autor-Rolle im Prolog ihre Inspirationsbitte eingeleitet:

*MARIA müter chvniginne | al der werld löferinne | verlihe mir vrowe folhe  
 finne | Daz ich ditz p̄chelinf beginne | Da ich dich mvge loben inne | Da  
 von ich din genade gewinne | Vnd iefus dines kindes minne | Des hilf mir  
 weifev meistferinne (v. 1–8).<sup>3</sup>*

Die exponierte Position im Prolog und die Durchreimung der ersten acht Verse gegen das Prinzip des höfischen Paarreimverses<sup>4</sup> befördern ebenso wie die identische Reimresponson *chvniginne/kuniginne – löferinne/löferinne*, (v. 1 f./v. 10203 f.) und die markanten Bezeichnungen Mariens die Assoziation zwischen beiden Textpassagen. Philipps ‚Marienleben‘ schlägt hier einen intratextuellen Bogen<sup>5</sup> zwischen dem Anfang des Textes und der ausklingenden Erzählung; es bezieht die Inspirationsbitte des Autor-Ichs auf die Himmelfahrt Mariens und führt durch Variation (v. 10204: *Ifraheliph löferinne* vs. v. 2: *al der werld löferinne*) den Perspektivwechsel des mittelalterlich-christlichen Autors gegenüber den antiken jüdischen Figuren des Textes vor.<sup>6</sup>

Es handelt sich allerdings nicht um die erste sprachliche Wiederholungsstruktur, die der Text anbietet. So klagt etwa Maria im Mater-dolorosa-Motiv:

*Vnd du mit grözer marter nôt | Liebez kint von mir scheidest | Vnd den  
 pittern töt von nu lidest | We mir we wer tötet mich | We mir we wer  
 tötet mich | Daz ih mit minem liebn kinde | Min vil armez leben ende |  
 owe mir ð grözn nôt | We mir we mines kindes töt | Daz ih den müz an  
 fehn | Vñ vor mir an ein † hehen | Vnd mag im doch gehelfen niht | Da*

---

Handschrift; vgl. zur Transkription <https://github.com/Transcriptions-Middle-High-German/Phillip-von-Seitz-Marienleben> (Zugriff: 08.11.2024). Vgl. hierzu unten 1.1. Alle Übersetzungen stammen von den Verfasser:innen.

- 2 Vgl. zur Himmelfahrt Mariens im ‚Marienleben‘ Ostermann 2019, S. 88–100.
- 3 „Maria, Mutter und Königin, Erlöserin der ganzen Welt, verleihe mir, Herrin, solche literarische Kunstfertigkeit, dass ich dieses Büchlein recht anfangen, mit dem ich dich zu preisen vermag, wodurch ich deine und Jesu, deines Kindes, Gnade gewinne. Inspiriere mich bei dieser Aufgabe, allwissende Meisterin.“
- 4 Vgl. Rückert 1853, S. 325–327.
- 5 Vgl. Lobsien 1995, S. 15: „Die Wiederholung ist eine intratextuelle Abbildung. Das als ‚(bloße)‘ Wiederholung qualifizierte Element verweist auf ein früheres, das in ihm identisch dasselbe oder doch als ein gleiches abgebildet ist.“
- 6 Vgl. zum Prolog des ‚Marienlebens‘ Gay-Canton 2009, S. 55–57.

von minem herzen we gefchiht | We mir we der mich nv vienge | Vnd an  
 daz selbe krevze hienge | Daz ih mit minem liebν kinde | Min vil armez  
 lebn ende (v. 7755–7771).<sup>7</sup>

Wörtliche Dopplungen (v. 7758 f. und v. 7760 f./v. 7770 f.), anaphorische Versstrukturen (v. 7758 f./7763/7768) oder leicht variierte Verse mit identischen Reimen (v. 7755/7762) bzw. einem Polyptoton am Versschluss (v. 7765/7769) erzeugen hohe sprachliche Kohärenz.<sup>8</sup> Wie im ersten Textbeispiel scheint dies semantisiert: Das gedoppelte Leid-Motiv der Passion Christi und des Mit-Leidens Mariens spiegelt sich in den sprachlichen Dopplungen; der gedoppelte Klageruf *We mir we wer tötet mich* (v. 7758 f.) ist durch den hier nicht durchgeführten Versumbruch auch im Layout der Handschrift als Dopplung hervorgehoben; wieder scheint die Variation innerhalb der Wiederholung interpretierbar, in diesem Fall sogar bis auf die graphematische bzw. hier piktographische Anlage der Handschrift.<sup>9</sup>

Beide Beispiele legen zwei Fragen nahe: 1. Stellen die sowohl den Text umschließenden als auch in unmittelbarem Kontakt stehenden sprachlichen Wiederholungen ein Strukturmerkmal des ‚Marienlebens‘ Philipps dar? 2. Handelt es sich hierbei um ein Spezifikum des Textes oder sind solche Strukturen typisch, zum Beispiel für geistliche Texte der Zeit? Für diese Fragestellungen wird im Folgenden ein Verfahren vorgeführt, mit dem das ‚Marienleben‘ (und potentiell jeder andere paargereimte Text)<sup>10</sup> systematisch auf Wiederholungs- bzw.

7 „Und Du, geliebtes Kind, scheidest mit großen Folterqualen von mir und erleidest jetzt den bitteren Tod. Weh mir, weh, wer tötet mich? Weh mir, weh, wer tötet mich, damit ich gemeinsam mit meinem sehr geliebten Kind mein Leben beenden möge? Weh mir, mich bekümmert das große Leid! Weh mir, mich bekümmert der Tod meines Kindes! Dass ich den mit ansehen muss und dass es vor mir an ein Kreuz gehängt wird und ich ihm doch nicht helfen kann, das stürzt mein Herz in Trauer. Weh mir, weh, wenn doch nur irgendwer mich festnehmen und an dasselbe Kreuz hängen würde, damit ich zusammen mit meinem geliebten Kind mein von nun an gänzlich armseliges Leben beenden könnte.“

8 Vgl. Flecken-Büttner 2011, S. 7: „Kohärenzstiftung durch Wiederholungs- und Variationsrelationen gehört zu den narrativen Universalien.“ Vgl. auch Hausmann 2010, S. 82: „Für den mentalen Wahrnehmungsapparat des Menschen, der wesentlich auf dem (Re)konstruieren von Mustern beruht, ist die Kategorie ‚Wiederholung‘ von elementarer Bedeutung: Ohne die Vorstellung, dass sich ‚etwas‘ in der Zeit (aber auch im Raum) wiederholt, gäbe es ‚für uns‘ weder synchrone (räumliche) noch diachrone (zeitliche) Muster. Dabei läuft die Wahrnehmung von Wiederholungen auf eine Komplexitätsreduktion hinaus (und das ist ja auch ihre Funktion): Tatsächlich nämlich wiederholt sich kein Ereignis und keine Situation ‚vollständig‘, vielmehr kann es sich immer nur um Ähnlichkeit handeln.“

9 Dem ‚tatsächlichen‘ Kreuz Christi, das in der Passionspassage in dieser Handschrift mit dem Kreuz-Piktogramm dargestellt ist, steht das von Maria ‚sprachlich‘ imaginierte Kreuz entgegen, das als sprachliches Lexem realisiert ist.

10 Vgl. Schwarzbach-Dobson 2022, S. 351: „Ein grundlegendes Prinzip des Erzählens ist die Wiederholung“; vgl. Lieb 2001, S. 507: „[D]ie Wiederholung ist – vielleicht in ausgezeichneter

Ähnlichkeitsstrukturen untersucht werden kann. Der hierdurch gewonnene strukturelle Befund wird im Anschluss gattungssystematisch kontextualisiert.

## 1 Datenerhebung

### 1.1 Textgrundlage: Das ‚Marienleben‘ in der Überlieferung der Handschrift Pommersfelden, Gräfl. Schönbornsche Schlossbibl., Cod. 46 (2797)

Das ‚Marienleben‘ Philipps von Seitz entstand Anfang des 14. Jahrhunderts.<sup>11</sup> Heute sind noch mehr als 100 Handschriften aus zwei Jahrhunderten und fast allen Sprachräumen des Deutschen erhalten. Es handelt sich ausweislich der Überlieferung um den am weitesten verbreiteten Versroman des Mittelalters in deutscher Sprache;<sup>12</sup> vermutlich sind „von keiner anderen deutschen Dichtung des 14. und 15. Jahrhunderts so viele Menschen erreicht worden“<sup>13</sup>. Bereits seit den Arbeiten Kurt GÄRTNERS ist allerdings bekannt, dass die noch immer einschlägige Edition RÜCKERTS (1853) keine valide Textgrundlage bietet.<sup>14</sup> Ebenfalls bekannt ist, dass die Handschrift Pommersfelden, Gräfl. Schönbornsche Schlossbibl., Cod. 46 (2797), im Folgenden Hs. P, „[v]on allen Hss. [...] den autornächsten Text [überliefert]; in ihr ist Philipp am ehesten faßbar, und alle künftigen Versuche, Philipps Sprache einschließlich seiner Reimsprache näher zu bestimmen, müssen von einer gründlichen Analyse der Sprache von P ausgehen“<sup>15</sup>. Damit steht die historische Textgrundlage für Untersuchungen fest,<sup>16</sup> die auf syntagmatische Strukturen, also auf die Sprache Philipps zielen;<sup>17</sup> Handschrift P war der Forschung bislang jedoch nicht zugänglich.

---

Weise im Mittelalter – auch ein Grundprinzip aller Formen kultureller Prozesse.“ So auch Roßbach u. Schrott 2023, S. 3.

11 Vgl. Gärtner 1978, S. 1.

12 Vgl. Gärtner 2014, S. 142; Masser 1987, S. 145, spricht von „eine[r] der meistgelesenen deutschen Dichtungen nicht nur des 14. und 15. Jahrhunderts, sondern wohl überhaupt des Mittelalters“.

13 Gärtner 1978, S. 1.

14 Vgl. ebd., S. 14: „Das Fazit meiner Kritik [...] kann nach allem nur lauten: unhistorisch, unkritisch, unzuverlässig und für wissenschaftliche Zwecke unbrauchbar.“ Eine Neuedition ist angekündigt, vgl. Gärtner 1994, S. 33–43.

15 Gärtner 1978, S. 370. Vgl. zu dieser Handschrift ebd., S. 160–164.

16 Vgl. Gärtner 1994.

17 Gerade die Sprache des ‚Marienlebens‘ hat zur Abwertung des Textes in der modernen Forschung geführt, während die zeitgenössischen Rezipienten diese offenbar schätzten; „vielmehr scheinen der einfache, an der Bibel geschulte Stil und die prosanahe Sprache des

## 1.2 OCR4all

Für diesen Beitrag konnte Handschrift P zunächst digitalisiert und im Anschluss mittels der Open-Source-Software OCR4all<sup>18</sup> automatisch transkribiert werden. HTR-Transkriptionen (*Handwritten Text Recognition*) erfolgen grundsätzlich in drei Schritten: Nach der Vorverarbeitung<sup>19</sup> schließt sich erstens die Layoutanalyse an, bei der der Text von anderen Elementen getrennt wird und die Zeilen im Bild identifiziert werden. Auf dieser Grundlage wird zweitens eine erste Transkription weniger Seiten mit Hilfe eines gemischten Grundmodells<sup>20</sup> durchgeführt und deren Qualität evaluiert. Handschrift P wurde vom aktuellen Grundmodell mit einer CER (*Character Error Rate*) von 3,45% erkannt.<sup>21</sup> Die hierbei anfallenden Transkriptionsdaten werden drittens für das Training eines schreiberspezifischen Modells (SSL) verwendet, wodurch ein auf die konkrete Handschrift angepasstes Transkriptionsmodell entsteht. Wenn dessen Qualität nicht ausreicht, wird dieser Prozess wiederholt.<sup>22</sup> Die Transkription von Handschrift P erreichte mit dem ersten SSL eine CER von 0,65% (1 Fehler auf 150 Zeichen). Händisch korrigiert wurden insgesamt 93 Zeilen. Mit diesem Modell konnten im Anschluss 267 Seiten vollautomatisch transkribiert werden, der automatische Textherstellungsprozess dauerte ca. 30 Minuten. Der manuelle Aufwand lag bei ca. einer Stunde. Mit Hilfe von HTR-Verfahren ist es daher auch jenseits von größeren Forschungsprojekten möglich, auf bisher unerschlossene Quellen zurückzugreifen. Hierbei muss zwar eine gewisse Fehlertoleranz in Kauf genommen werden, was angesichts der massiven Zeitersparnis allerdings vertretbar scheint.

---

Autors die literarischen Vorzüge der ursprünglichen Fassung des Werkes auszumachen und auch seinen Erfolg zu erklären“ (Gärtner 2014, S. 143).

18 Vgl. <https://www.ocr4all.org/> (Zugriff: 08.11.2024).

19 Enthält z.B. Operationen wie das Geradestellen des Scans oder die Umwandlung in ein Schwarzweißbild.

20 Vgl. Reul, Tomasek, Langhanki u. a. 2022; vgl. zu den verwendeten Modellen <https://www.germanistik.uni-wuerzburg.de/handschriften/> (Zugriff: 08.11.2024).

21 Die CER wird durch Abgleich zwischen der automatischen und der händisch erstellten Transkription gebildet; sie ergibt sich aus dem Verhältnis von fehlerhaften Zeichen zu allen Zeichen.

22 Vgl. zu diesem iterativen Prozess Reul, Tomasek, Langhanki u. a. 2022.

## 2 Distanzmaße und Methode

### 2.1 Distanzmaße

Für die automatische Erkennung von Wiederholungs- bzw. Ähnlichkeitsstrukturen wäre ein Zugriff auf Lexemebene naheliegend. Mit *Word-Embedding*-Verfahren liegen zuverlässige, trainingsbasierte Algorithmen vor, die wortähnliche Verse automatisch erfassen könnten. Allerdings steht der automatisierten Analyse von mittelhochdeutschen/frühneuhochdeutschen Lexemen der nicht normalisierte Sprachstand entgegen, durch den auch innerhalb des gleichen Textes erhebliche orthographische, dialektale oder auf diachronem Sprachwandel beruhende Varianten einkalkuliert werden müssen.<sup>23</sup> Ein *Word-Embedding*-Modell für diplomatische Transkriptionen von mittelhochdeutschen bzw. frühneuhochdeutschen Handschriften müsste auf einem ausreichend großen Trainingskorpus aufbauen, das die Varianten des nicht normalisierten Sprachstands enthält. Ein solches Trainingssetting liegt bisher nicht vor.<sup>24</sup> Zudem sind die bestehenden Korpora quantitativ relativ schwach, wodurch mit Fehlzuordnungen zu rechnen ist. Daher wurden für diesen Beitrag andere Distanzmaße eingesetzt:

1. Levenshtein-Distanzen (LevD): Levenshtein-Distanzen zählen Operationen, die notwendig sind, um eine Zeichenkette *x* in eine Zeichenkette *y* zu transformieren. Hierbei werden Löschen-, Einsetzungs- und Ersetzungsoperationen gezählt. Der daraus resultierende Distanzwert ist eine numerische Zahl; der kleinstmögliche Wert ist null (vollständige Übereinstimmung), der größtmögliche Wert entspricht der Länge der größeren Zeichenkette. Dieses basale Verfahren hat den Vorteil, dass es auf beliebige Zeichenketten anwendbar ist und damit weder durch vor allem graphematische Varianten beeinflusst wird noch trainiert werden muss. LevD-Werte bilden daher das Grundgerüst der nachfolgenden Analysen. Ein entscheidender Nachteil ist allerdings die sehr geringe Ausdifferenzierung der Distanzwerte.
2. Normalisierte Levenshtein-Distanzen (LevN): LevD-Werte sind prinzipiell kaum untereinander vergleichbar, da sie unmittelbar von der Länge der jeweiligen Zeichenketten abhängen. Daher kann es sinnvoll sein, die Zahl

<sup>23</sup> Vgl. für Handschrift P z.B. die Schreibung *da3* (807) vs. *daz* (400); *fprach* (312) vs. *fprah* (17); *iefus* (283) vs. *iefuf* (18) vs. *iefus* (9) und die flektierten Formen *iefum* (65) bzw. *iefu* (61); *hüs* (30) vs. *haus* (22) vs. *hus* (1) vs. *hauf* (7) bzw. flektiert *haufe* (4) etc.

<sup>24</sup> Vgl. Tomasek, Reul u. Wehner 2022. Die Alternative bestünde in einer automatischen Normalisierung des Untersuchungstextes. Hierzu gibt es bereits Methoden wie *Part-of-Speech Tagger* o. ä. Allerdings konnten für die Handschrift P mit einer Erkennungsquote < 75 % keine suffizienten Ergebnisse erreicht werden.

der Operationen prozentual zur Zahl der maximal möglichen Operationen in Beziehung zu setzen, um einen Prozentwert der veränderten Zeichen zu ermitteln (LevN).<sup>25</sup> Diese Werte sind wesentlich ausdifferenzierter als LevD und daher gerade bei der Festlegung der n-ähnlichsten Verse produktiv. Die Normalisierung verzerrt allerdings die Distanzmaße, weil zum Beispiel ein unnormalisierter LevD = 1 (eine Ersetzungsoperation) bei einer Zeichenkette von 10 Zeichen einen normalisierten LevN = 0,1 ergibt, bei einer Zeichenkette von 20 Zeichen hingegen LevN = 0,2 etc. Daher werden im Folgenden immer beide Lev-Werte betrachtet.

3. Gewichtete Levenshtein-Distanzen (LevG): Aufgrund des nicht normalisierten Sprachstands sind viele von Lev-Werten erfasste Varianten unterhalb der phonetisch relevanten Sprachebene angesiedelt. Das kann relevant sein, hat aber oft den Effekt, dass eine höhere Varianz angezeigt wird, als es für die jeweilige Fragestellung sinnvoll ist. Ein typisches Beispiel wäre der Austausch von ‚v‘ und ‚f‘, bei dem zu erwarten ist, dass es sich um eine Graphemvariante handelt. Da mittelhochdeutsch ‚v‘ aber auch [u] bzw. [w] bezeichnet, kann auch eine phonetisch relevante Variante vorliegen. Bei LevG können Ersetzungsoperationen dann schwächer gewichtet werden, wenn sie mutmaßlich wenig Aussagekraft für das intendierte Ergebnis haben, aber nicht sicher als Variante auszuschließen sind. Im Fall der Handschrift P wurden daher zunächst alle vorkommenden Grapheme automatisch identifiziert und dann gewichtet; es werden Gewichtungen zwischen 0 (Austausch zwischen zwei Zeichen wurde nicht gezählt), 0,5 (Austausch zwischen zwei Zeichen wurde nur als halbe Operation gezählt) und 1 (Austausch zwischen zwei Zeichen wurde regulär gezählt) angewendet.
4. *Jaccard-Distance* (JacD): Lev-Werte messen generell von links nach rechts. Damit sind sie besonders sensitiv für syntaktische Umstellungen. Hierdurch können identische oder sehr ähnliche Verse mit hoher LevD gemessen werden, obwohl paradigmatisch die gleichen Lexeme, allerdings in veränderter syntagmatischer Stellung, vorliegen. So würde im folgenden Beispiel die Umstellung *al der werld löferinne* (v. 2) zu *\*löferinne al der werld\** eine LevD 19/LevN 0,83 (statt LevD 11/LevN 0,47) ergeben. Damit solche Verspaare dennoch erkannt werden, wurde auf Jaccard-Distanzen (JacD) als Maß für Versähnlichkeiten zurückgegriffen. Dieses Distanzmaß setzt die Schnittmenge (alle beiden Versen gemeinsamen Objekte) in Relation zur Vereinigungsmenge (alle Objekte aus beiden Versen): Von der

---

<sup>25</sup> Im Folgenden mit einem LevN-Wert zwischen 0 und 1 wiedergegeben, wobei 1 = 100% und 0 = 0% der maximal möglichen Ersetzungsfunktionen angibt.

Vereinigungsmenge wird die Schnittmenge subtrahiert und dieser Wert durch die Vereinigungsmenge dividiert. Wenn beide Zeichenketten identisch sind, ist der JacD-Wert dementsprechend 0, der höchste mögliche Wert ist 1 (keine Schnittmenge). Dieses Verfahren hat allerdings zwei Defizite: Einerseits wird nur das Vorkommen, nicht aber die Häufigkeit der Objekte gemessen. Deshalb kann eine Differenz, die sich daraus ergibt, dass ein identisches Objekt in einer Zeichenkette zweimal, in der Vergleichs-Zeichenkette aber nur einmal vorliegt, nicht erfasst werden. Die Wahrscheinlichkeit für das Auftreten dieses Effekts wird durch die Verwendung von *strings* (s. u.) verstärkt. Andererseits ist dieses Distanzmaß zwar ausdifferenzierter als die LevD, aber weniger spezifisch als LevN, wodurch es relativ häufig zu Dubletten gleichwertiger Distanzwerte kommt. In diesen Fällen sind die LevN-Werte stets aussagekräftiger als die JacD-Werte.

Aufgrund der hohen Varianz auf Lexemebene (vgl. Anm. 23) ist die Schnittmenge zwischen zwei Versen für JacD oft nicht suffizient messbar, weil eigentlich identische Lexeme als voneinander abweichend klassifiziert werden. Um diesen Effekt auszugleichen, wurden die jeweiligen Lexeme unter Verwendung von N-Grammen in *strings* mit je 2 Graphemen zerlegt, das heißt, dass zum Beispiel das mittelhochdeutsche Lexem *muoter* in sechs *strings* (mu-uo-ot-te-er- r+Leerzeichen) geteilt wird, wodurch bei der Variante mittelhochdeutsch *muter* (mu-ut-te-er-r+Leerzeichen) immer noch vier (von fünf) *strings* identisch sind. Hierdurch können graphematisch oder orthographisch sowie dialektal bedingte Varianten besser aufgefangen werden. Dies führt allerdings dazu, dass in einigen Konstellationen apokopierte Endsilben (vor allem -e) nicht als Distanz gemessen werden, weil der in einem Vers fehlende *string* an anderer Position im Vers vorkommt.<sup>26</sup> Aufgrund dieser Unschärfe werden die JacD im Folgenden nur in Relation zu LevN verwendet.

## 2.2 Algorithmus

Mit Hilfe dieser Distanzmaße wurden für jeden Vers des ‚Marienlebens‘ die *n*-ähnlichsten Verse innerhalb des Textes automatisch bestimmt.<sup>27</sup> Hierfür wurden die

---

<sup>26</sup> Apokopiertes ‚-e‘ kann nicht gemessen werden, wenn das vorletzte Zeichen plus ‚-e‘ an anderer Position des Verses bereits vorkam.

<sup>27</sup> Vgl. für den verwendeten Algorithmus <https://github.com/Toolbox-Middle-High-German/Automatic-Intratextual-Coherence-Analysis> (Zugriff: 08.11.2024).

HTR-Daten zunächst vorverarbeitet.<sup>28</sup> Anschließend erfolgte die Berechnung der Distanzen aller 10.440 Verse der Handschrift zueinander, also  $4 \times 10.440^2 = 435.974.400$  Rechenvorgänge. Die Matches wurden mit Versnummer und Distanzwerten tabellarisch ausgegeben. Es ist evident, dass ein solches Setting händisch kaum zu bewerkstelligen wäre, es aber eine prototypische Aufgabe für einen Algorithmus darstellt. Allerdings zeigten sich Grenzen der Operationalität, da die hohe Zahl an Berechnungen selbst bei Parallelisierung der Rechenvorgänge zu Laufzeiten von mehreren Stunden führte.

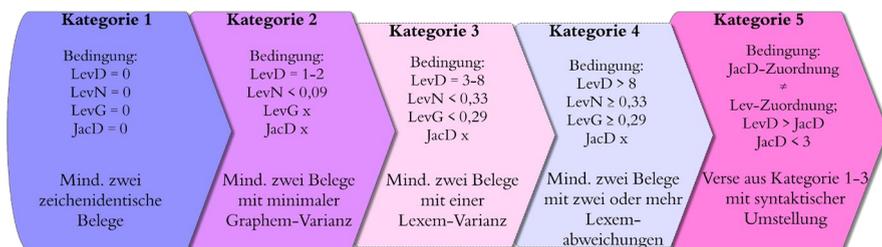
### 2.3 Grenzwerte

Mit diesem Verfahren wurden fünf Kategorien für Ähnlichkeitsbeziehungen abgeleitet, die für die Fragestellung dieses Beitrags relevant sind. Die Kategorien wurden jeweils nur für das ähnlichste Match zwischen zwei Versen, also  $n = 1$ , gemessen (vgl. Abb. 1). Ab Kategorie 3 zeigen sich die oben beschriebenen Verzerrungseffekte: Verse mit vielen Zeichen werden von den LevD höher eingestuft als von den LevN, bei kurzen Versen verhält es sich umgekehrt. Daher werden immer beide Werte als Klassifikator genutzt. Die Zuordnung ist abgestuft: Je höher die Distanzwerte, umso größer ist die Wahrscheinlichkeit, dass zwei (oder mehrere) Lexeme abweichen. Ab LevD 5/LevN 0,16 sind die Unterschiede oft so hoch, dass sich stärkere Bedeutungsdifferenzen zwischen den Versen ergeben. Während in Kategorie 1 und 2 also Spiegelungen von Versen im Vordergrund stehen, tritt ab dieser Distanz die Variation bei der Verswiederholung stärker hervor. Die Grenzwerte zu Kategorie 4 sind bisweilen nicht trennscharf. Ab Kategorie 4 ist zu erwarten, dass einige nicht korrekt identifizierte Verse zugeordnet sind, die durch viele sprachliche Varianten unterhalb der Lexemebene in den LevD-Werten zu hoch eingemessen werden. Die hier angesetzte Grenze ist also eher restriktiv und erfasst potentiell zu wenige Verszuweisungen.

Die gewichteten LevD-Werte LevG erzielten zwar deutlich andere Verszuordnungen für das ähnlichste Match als die LevN-Werte (2.275 = 21,4 % der Verspaare wurden anders als durch LevN gebildet); diese waren allerdings nie relevant für die ersten beiden Kategorien, sondern setzen bei LevN = 0,16 (d. i. LevG = 0,11) ein, also ab Kategorie 3 und hier auch bereits im höheren Distanzbereich. In dieser Gruppe wurden für 176 Verse die ähnlichsten Matches anders gebildet als durch LevN, in der Regel tauschten aber gegenüber LevN nur das ähnlichste mit dem zweitähnlichsten Match die Position. Da in diesem Bereich ohnehin verstärkt Lexemvarianten die Ähnlichkeitseffekte überlagern, ist eine verbesserte Zuordnung durch

---

<sup>28</sup> Entfernung nicht gewünschter Metadaten; Herstellung der Zeilenumbrüche; Korrektur der am Zeilenanfang ausgerückten Erstbuchstaben.



**Abb. 1** | Kategorien für Vers-Ähnlichkeitsverhältnisse im ‚Marienleben‘. Graphik: Stefan Tomasek, Kiara Hart, Franziska Schulte und Christian Reul.

LevG hier nicht erkennbar. Die ohnehin unscharfe Grenze gegenüber Kategorie 4 konnte ebenfalls nicht trennschärfer modelliert werden, da zwar die Werte insgesamt kleiner, aber mit der Perspektive auf Lexemabweichungen nicht konziser wurden. Der LevG-Grenzwert zu Gruppe 4 liegt bei  $\text{LevG} < 0,29$ .

In Kategorie 5 wurden mit JacD im (vermeintlich) totalidentischen Bereich (fälschlich) drei zusätzliche Verse eingemessen, die sich aus dem Prinzip der Vereinigungsmenge ergeben (vgl. Anm. 26).<sup>29</sup> Generell weisen die JacD zwar auch deutlich andere Ergebnisse bei der Identifizierung der ähnlichsten Verspaare auf (7.526 Verspaare weichen ab). Das liegt allerdings am relativ unspezifischen Distanzmaß, durch das sich häufig mehrere Verspaare mit identischer JacD ergeben und hier dann eine an der Versreihenfolge orientierte Zuordnung erfolgt. JacD-Werte sind also für die Fragestellung dieses Beitrags nur in Relation zu Lev-Werten produktiv, obwohl sie aufgrund ihrer zwei vollständig unterschiedlichen Wertesysteme eigentlich nicht miteinander verrechenbar sind. Allerdings hat sich gezeigt, dass die LevN-Werte bei der Identifizierung der ähnlichsten Verspaare im Durchschnitt kleiner als JacD-Werte sind und die LevN-Verspaare immer dann die deutlich plausibleren Treffer ergeben, wenn  $\text{LevN} < \text{JacD}$  ist. Wenn umgekehrt  $\text{LevN} > \text{JacD}$  bei der Berechnung des ähnlichsten Verses ausfällt, liegt das fast immer an einer veränderten Satzstellung. Bei den so identifizierten Versen mit veränderter Satzstellung sind Grenzwerte deutlich schlechter zu definieren als bei den Lev-Werten, da die veränderte Syntax häufig darauf zurückgeht, dass Satzteile im vorherigen oder folgenden Vers stehen, Subjekt-Objekt-Beziehungen vertauscht oder Satzarten verändert sind. Daher treten bereits bei sehr niedrigen JacD Unterschieden auf Lexemebene auf,<sup>30</sup> andersherum werten auch höhere JacD

<sup>29</sup> Vgl. v. 229 (*Dv̇ der engel dife red*) zu v. 269 (*Dv̇ der engel dife rede*); v. 1935 (*Si gewan nie mannes gemein*) zu v. 3574 (*Si gewan nie mannes gemeine*); v. 9293 (*Vnd die krumpen rehte gente*) zu v. 9757 (*Vnd die krumpen rehte gent*).

<sup>30</sup> Vgl. z.B. v. 9973 (*Von dir ift al div werlt erlöft*), dem LevN als ähnlichsten Treffer *Von dir wirt al dev werlt erlöft* (v. 1616) zuordnet, während JacD v. 6430 (*Von dir al div werlt erlöft*)

auf Lexemebene Verse als identisch, bei denen die Satzstellung getauscht ist.<sup>31</sup> Insofern fehlt eine stringente Trennung zwischen Spiegelungen und Variationen, wie sie durch die Lev-Werte gezeigt werden. Dennoch liefern diese Werte zusätzliche Treffer, die durch die Lev-Werte nicht erfasst werden können. So befindet sich etwa das Verspaar v. 7732 (*Dv wære min khneht du wære min herre*) mit v. 1434 (*Dv piſt min herre du piſt min khneht*), das durch JacD als ähnlichster Treffer identifiziert wird, aufgrund der Variante des Verbs und der Umstellung der Satzteile durch LevN nicht unter den zehn ähnlichsten Verspaaren und ist folglich nicht in Kategorie 1 bis 3 einsortiert, es wäre daher auch nicht ins Untersuchungssetting eingegangen. Ab JacD > 0,3 werden die Ähnlichkeiten zwischen den Versen deutlich geringer, so dass sie ab hier für diesen Beitrag nicht mehr herangezogen wurden. Einen klaren Grenzwert gibt es allerdings nicht.

### 3 Quantitative Analyse des ‚Marienlebens‘

*Kategorie 1:* Handschrift P enthält 208 Verse, die mindestens einen totalidentischen Vers an anderer Position aufweisen (von 10.618 Versen einschließlich der Zwischenüberschriften).<sup>32</sup> Zeichenidentisch, aber durch die geänderte Satzstellung nicht von den Lev-Werten erfasst, sind drei weitere, durch JacD identifizierte Verse.<sup>33</sup> Damit enthält das ‚Marienleben‘ 208/212 zeichengenau identische Verswiederholungen (= 1,96 bis 2,0 % des Gesamttextes).<sup>34</sup> Dass so viele identische

---

vorschlägt, auf den in v. 6431 das flektierte Verb *Wirt* [...] folgt. Tatsächlich ist dies (aufgrund der Variante *div* vs. *dev*) der ähnlichere Treffer, aber nur, wenn man die Verbenstellung und das Enjambement auflöst.

- 31 Vgl. z. B. das Verspaar v. 238 (*Sælich ift din fel und auh din lip*) mit v. 1661 (*Din fele ift sælich und auch din lip*), das mit einer hohen JacD (0,25) als ähnlichster Treffer eingemessen wird. Auf Lexemebene sind beide Verse identisch, die Unterschiede liegen in der Wortstellung. Die LevN-Zuordnung erfasst dieses Verspaar erst als drittähnlichsten Treffer nach der Zuordnung von v. 1479 (*Rein ift fin fele vñ khevfch fin lip*) und v. 1037 (*Min fele ift fin und auch min lip*); das Verspaar mit v. 1661 wäre mit LevN = 0,44 aus der Kategorie 3 gefallen und daher ohne die JacD nicht in das Untersuchungssetting eingegangen.
- 32 Einen Grenzfall stellt das Verspaar v. 8516 (*Maria sprach min herre min fun*) zu v. 5276 (*Maria sprach fun herre min*) dar, das aufgrund der Lev-Werte nicht als identisch identifiziert werden kann, aus der Perspektive der Vereinigungsmenge jedoch als ein identisches Verspaar erkannt wird, da hier das fehlende gedoppelte *min* und die geänderte Wortstellung keine Rolle spielen.
- 33 Vgl. v. 4398 (*Iefus sprach dv müter min*) zu v. 5161 (*Iefus dv sprach müter min*); v. 7152 (*An fin wange in harte flügen*) zu v. 8865 (*An fin wange flügen in harte*).
- 34 Durchschnittlich tritt ein totalidentischer Vers alle 224 Zeilen auf. Handschrift P ist mit 29 bis 31 Zeilen pro Seite angelegt, wobei bisweilen zwei Verse mit Reimpunkt nacheinander eingetragen sind. Damit liegen also im Durchschnitt ca. sieben Seiten zwischen den identischen Versen. Dass diese direkt zueinander in Kontakt stehen und sich die identische Anlage aus

Verse auftreten, spricht zum einen für die einheitliche Anlage der Vorlage von Handschrift P bzw. den sorgfältigen Schreibprozess des P-Schreibers. Bedenkt man aber die Fülle an Variationsmöglichkeiten, die der nicht normalisierte Sprachstand geboten hätte (vgl. Anm. 23), ist dieser Befund überraschend.

*Kategorie 2:* Hierher wurden 282 Verspaare sortiert (= 2,66 % des Gesamttextes), davon wurden zwei nur durch die LevN-Werte und 21 nur durch die LevD-Werte zugeordnet.<sup>35</sup> Es zeigt sich also der beschriebene Verzerrungseffekt, der sich durch Messungen aus beiden Perspektiven auffangen lässt. Die Unterschiede zwischen den Versen sind fast nie bedeutungsrelevant. In 13 Fällen sind die gemessenen Kleinstvarianten das Ergebnis von verändertem Kasus oder Genus,<sup>36</sup> zweimal bezeichnen sie bedeutungsähnliche Lexeme.<sup>37</sup> In elf Fällen entstehen durch sie allerdings neue Bedeutungen (3,9 % der Gruppe; eigentlich Kategorie 3).<sup>38</sup> In die Kategorien 1 und 2 wurden damit 471 bis 500 Verspaare eingeordnet (ohne veränderte Satzstellungen und ohne Fehlzuordnungen), das entspricht 4,44 bis 4,7 % des Gesamttextes.

*Kategorie 3:* Hier wurden einerseits sehr ähnliche Verspaare wie v. 523 (*Nv fagt uns dev fchrift fvr war*) zu v. 861 (*Da3 fagt uns dev fchrift fvr war*), andererseits aber auch variantenreichere Verspaare wie v. 9363 (*wie fÿze und auch wie rehte fi wære*) zu v. 9348 (*Wie heilich und auch wie güt fi wære*) etc. eingruppiert (insgesamt 2.492 Verse = 23,47 % des Gesamttextes); 1.315 Verspaare wurden von LevD und LevN gemeinsam eingruppiert, 41 Verspaare nur über LevD, 1.136 Verse nur über LevN. Die Grenze zu Kategorie 2 ist unspezifisch: Im unteren Bereich (LevN 0,09–0,11) weisen von 75 Verspaaren 31 keine Lexemvarianten auf (eigentlich Kategorie 2). Bei veränderten Grenzwerten wären aber umgekehrt 44 Verspaare fälschlich Gruppe 2 zugeordnet worden. Das gleiche Problem zeigt sich in der Abgrenzung zu Kategorie 4. Präzise Grenzwerte zu definieren, stellt ein Defizit der gewählten

---

einem unmittelbaren Kopiervorgang erklären lassen könnte, ist selten: Auf fol. 100r stehen zwei identische Verse in direktem Kontakt, hier entgegen der grundsätzlichen Anlage der Handschrift in einer Zeile mit Reimpunkt (v. 7758f.); ansonsten stehen nur acht Verspaare auf der gleichen Seite (vgl. fol. 78r [v. 6036/6044], 34v [v. 2604/2612], 125r [v. 9837/9847], 100r [v. 7760/7770], 15r [v. 1059/1067], 31v [v. 2364/2378], 109v [v. 8538/8558], 38v [v. 2912/2944]. Ein Paar findet sich im unmittelbaren Seitenwechsel (vgl. 42r/42v [v. 3190/v. 3234]).

35 Erwartungsgemäß handelt es sich bei den beiden nur durch LevN erfassten Verspaaren um Verse mit überdurchschnittlich vielen Zeichen (z. B. die Zuordnung v. 3455 [*Marien di fÿzen und auch di reine*] zu v. 719 [*Marien die fÿzen und auch die reinen*], während die nur durch LevD zugewiesenen Verspaare kurze Verse sind (z. B. die Zuordnung v. 6329 [*In ð stat ierufalem*] zu v. 8011 [*In die stat ierufalem*]).

36 Vgl. *ð* vs. *di*; *der* vs. *den*; *dem* vs. *der*; *da3* vs. *den*; *er* vs. *fi* (dreimal); *im* vs. *ir* (zweimal); *in* vs. *im*; *ir* vs. *dir*; *fingent* vs. *fungen*; *springent* vs. *springen*.

37 Vgl. *dÿ* vs. *dar*; *werden* vs. *wefen*.

38 Vgl. *din* vs. *fin* (d. i. das Verb *sîn*); *min* vs. *vil*; *dÿ* vs. *Da3*; *ir* vs. *ein*; *nv* vs. *in*; *Vmb* vs. *Vñ*; *wol* vs. *waz*; *giengen* vs. *viengen*; *klagen* vs. *vragen*; *fÿnden* vs. *wnden*; *Martha* vs. *Maria*.

Methode dar. Generell ist zu erwarten, dass diese Grenzwerte werkspezifisch sind und für andere Texte nachjustiert werden müssen.

Dennoch lässt sich mit dieser quantitativen Herangehensweise bereits die erste Fragestellung des Beitrags beantworten: Gut 4% aller Verse des ‚Marienlebens‘ haben mindestens einen Pendant-Vers, der entweder völlig identisch ist oder nur Abweichungen unterhalb der Lexemebene aufweist; knapp 24% des Textes bestehen aus Verspaaren mit hohen Ähnlichkeitseffekten. Ca. 28% des Textes sind also durch intratextuelle Wiederholungsstrukturen geprägt. Damit sind Versdopplungen ein entscheidendes Strukturmerkmal des ‚Marienlebens‘ Philipps von Seitz.

*Kategorie 5:* Die durch die JacD-Werte identifizierten Verspaare stellen einen Sonderfall dar. Insgesamt wurden 430 neue Verspaare erfasst, von denen 46 Varianten der Kategorie 2, die übrigen Varianten der Kategorie 3 darstellen. Hierdurch erhöht sich der Anteil der Verse mit Wiederholungsstrukturen auf 32% des Gesamttextes. Allerdings stellt natürlich die syntagmatische Veränderung selbst bereits eine Variation dar.

#### 4 Einordnung in Gattungszusammenhänge

Um diesen Befund einzuordnen, bleibt zu klären, ob es sich hierbei um ein Spezifikum des ‚Marienlebens‘ handelt oder ob klerikale Texte (zum Beispiel aufgrund ihres didaktischen Charakters oder ihrer Nähe zur Liturgie etc.) generell solche intratextuellen Verweisstrukturen enthalten.<sup>39</sup> Zur Beantwortung dieser Frage wurde der Algorithmus so angepasst, dass mehrere Textdateien gleichzeitig eingelesen wurden, für jede Zeile aus jeder Datei das ähnlichste Match zugeordnet und dann eine statistische Auswertung ausgegeben wurde. Aufgrund der besseren Vergleichbarkeit ist LevN das gewählte Distanzmaß.

Als Vergleichsdaten wurden zunächst alle in der ‚Mittelhochdeutschen Begriffsdatenbank‘ (MHDBDB) unter der Rubrik ‚Religiöse Versdichtungen‘ gelisteten Texte ausgewählt.<sup>40</sup> Die insgesamt 56 Texte dieses Korpus wurden durch eine händisch erstellte diplomatische Transkription der ‚Kindheit Jesu‘ Konrads von Fußesbrunnen (Wiener Hs. B) erweitert. Für das ‚Marienleben‘ stellt dieser Text einen naheliegenden Vergleichstext dar, da Philipp ihn mit einiger Wahrscheinlichkeit kannte, die ‚Kindheit Jesu‘ das ‚Marienleben‘ stellenweise beeinflusst hat

<sup>39</sup> Vgl. Schwarzbach-Dobson 2022, S. 351: „Erzählschemata oder Gattungstypologien setzen wiederkehrende Muster voraus, die gerade durch ihre iterative Verwendung Konstanz und Kohärenz suggerieren.“

<sup>40</sup> Vgl. <http://mhdbdb.sbg.ac.at:8000/mhdbdb/App?action=TextList> (Zugriff: 08.11.2024). Für die Fragestellung dieses Beitrags wurden die dort gelisteten Prosatexte aussortiert.

und beide Texte in der Überlieferungsgeschichte des ‚Marienlebens‘ teilweise kompiliert wurden.<sup>41</sup>

Da den Daten der MHDBDB meist normalisierte Editionen zugrunde liegen, ist der Vergleich mit HTR-basierten Transkriptionen nicht völlig kohärent, weil vor allem durch das rekonstruierte ‚Normalmittelhochdeutsch‘ der Editionen generell niedrigere Lev-Werte zu erwarten sind. Zudem dürften die für das ‚Marienleben‘ erarbeiteten Grenzwerte nicht eins zu eins übertragbar sein; auch hier gilt allerdings, dass kleinere Distanzwerte bei den Editionen zu niedrigeren Grenzwerten führen, sodass den ersten vier Kategorien eher mehr Verse zugeordnet werden als beim ‚Marienleben‘.

Einen ersten Überblick kann bereits der Mittelwert für alle besten Matches pro Vers liefern. Im ‚Marienleben‘ wurde für die ähnlichsten Verse ein Durchschnittswert von  $\emptyset$  LevN 0,37 ermittelt, der Durchschnitt im Vergleichskorpus ‚Religiöse Versdichtung‘ liegt bei  $\emptyset$  LevN 0,48.

Obwohl man davon ausgehen muss, dass die Wahrscheinlichkeit für LevN = 0 aufgrund der Normalisierung der Editionen wesentlich höher ist, weisen 33 der 57 Texte überhaupt keine totalidentischen Verspaare auf, 14 Texte enthalten weniger als halb so viele wie das ‚Marienleben‘ (< 1,0 % aller Verse vs. ‚Marienleben‘ = 2,0 %); beim naheliegendsten Vergleichstext, Konrads von Fußesbrunnen ‚Kindheit Jesu‘, sind nur 0,07 % aller Verse zeichenidentisch. Lediglich acht Texte enthalten totalidentische Verse in vergleichbarer Größenordnung wie das ‚Marienleben‘: Rudolf von Ems, ‚Der guote Gêrhart‘ (1,3 %), ‚Der ertrunkene Glöckner A‘ (1,18 %), Konrad von Würzburg, ‚Engelhard‘ (1,38 %), Konrad von Würzburg, ‚Pantaleon‘ (1,39 %), Rudolf von Ems, ‚Barlaam und Josaphat‘ (1,81 %), Konrad von Würzburg, ‚Alexius‘ (1,98 %), Reinbot von Durne, ‚Der heilige Georg‘ (2,49 %) und Konrad von Würzburg, ‚Silvester‘ (2,6 %). Signifikant mehr totalidentische Verse enthält nur die ‚Jüngere Judith‘ (3,68 %), der quantitative Ausreißer ist der ‚Münchner Oswald‘ (10,58 %).

In Kategorie 2 setzt sich diese Tendenz fort: Hierher wurden 2,66 % der Verspaare des ‚Marienlebens‘ einsortiert; 33 (von 57) Texten des Vergleichskorpus zeigen in unserem Setting gar keine Zuordnung zu dieser Kategorie. Außer dem ‚Münchner Oswald‘, der auch hier wieder einen Ausreißer darstellt (5,92 %), kommt kein Text an die Häufigkeit des ‚Marienlebens‘ heran; die ‚Kindheit Jesu‘ liegt hier bei 0,77 %. Nimmt man beide Kategorien zusammen, um die Lexem-identischen Verse zu vergleichen, steht das ‚Marienleben‘ mit 4,44 bis 4,70 % aller Verse zwischen dem ‚Heiligen Georg‘ (5,22 %) und der ‚Jüngeren Judith‘ (4,23 %) und wird nur vom ‚Münchner Oswald‘ deutlich übertroffen (16,5 %).

---

41 Vgl. Gärtner 2008, S. 407.

In der (allerdings in diesem Vergleich vermutlich noch unspezifischer anzusetzenden) Kategorie 3 wird das ‚Marienleben‘ (23,47 %) nur vom ‚Münchner Oswald‘ (32,07 %) und in diesem Fall dem ‚Barlaam und Josaphat‘ (28,06 %) übertroffen, in vergleichbarer Größenordnung weisen nur noch ‚Der guote Gêhart‘ (21,34 %), ‚Engelhard‘ (20,19 %) und der ‚Silvester‘ (20,03 %) Verspaare dieser Kategorie auf, die ‚Kindheit Jesu‘ liegt bei 4,29 %.

Dieser Vergleich lässt sich auf andere Gattungssysteme ausweiten: Im Vergleichskorpus ‚Heldendichtung‘, das ohnehin Texte enthält, die als verdächtig für schematisches Erzählen gelten, wird das ‚Marienleben‘ in Kategorie 1 und 2 vor allem von der Dietrichsepik deutlich übertroffen, wobei umgekehrt etwa der ‚Goldemar‘ überhaupt keine Wiederholungsstrukturen in dieser Kategorie enthält. Ähnlich verhält es sich interessanterweise auch bei Texten wie der ‚Kudrun‘, dem ‚Nibelungenlied‘ oder der ‚Klage‘, die hier Werte < 0,5 % aufweisen. Insgesamt zeigen von 32 Texten elf (Dietrichs-)Epen höhere Werte in den Kategorien 1 und 2 als das ‚Marienleben‘. Aus der Perspektive der intratextuellen Wiederholungsstrukturen ist Philipps Text mit dem ‚Rolandslied‘, ‚Herzog Ernst B‘ oder Strickers ‚Karl der Große‘ vergleichbar.<sup>42</sup> Das Vergleichskorpus ‚Artusdichtung‘ enthält erwartungsgemäß in Kategorie 1 und 2 deutlich niedrigere Durchschnittswerte als das ‚Heldenepik-Korpus‘. Dementsprechend wird hier das ‚Marienleben‘ nur von Wirnts von Gravenberg ‚Wigalois‘ und Texten des Pleiers übertroffen; letztere stellen die Ausreißer in diesem Korpus dar. Allerdings befindet sich Philipps Text in der Kategorie 1 (LevN = 0) mit 2,00 % aller Verse in einer ähnlichen Größenordnung wie Wolframs ‚Parzival‘ (1,75 %), Hartmanns ‚Erec‘ (1,84 %) oder Gottfrieds ‚Tristan‘ (2,08 %). Im Vergleichskorpus ‚Klassische Epen‘ (= Antikenromane) erreichen in der Kategorie 1 und 2 neben Herborts ‚Liet von Troye‘ wieder Rudolf von Ems ‚Alexander‘ und Konrad von Würzburg ‚Trojanerkrieg‘ höhere Werte als das ‚Marienleben‘, das in dieser Gruppe hinsichtlich der intratextuellen Referenz etwa dem ‚Straßburger Alexander‘ des Pfaffen Lambrechts entspricht, deutlich übertroffen wird es von Veldekes ‚Eneas‘ (8,04 %). Der Bereich der Kleinepik ist aufgrund der relativen Kürze der Texte schwer vergleichbar.

Die zweite Frage dieses Beitrags nach intratextuellen Wiederholungen als gattungsspezifischem Strukturmerkmal geistlicher Texte lässt sich also in doppelter Hinsicht verneinen: Weder weisen die Mehrzahl der klerikalen Texte vergleichbare Strukturen auf wie das ‚Marienleben‘ noch fehlt dies in anderen Gattungen. In der geistlichen Epik sind es offenbar die Texte bestimmter Autoren (Rudolf von Ems, Konrad von Würzburg, Reinbot von Durne), die sich ähnlich verhalten wie der Text Philipps. Da sich dies bei Rudolf von Ems und Konrad von Würzburg auch

---

<sup>42</sup> Die MHDBDB führt diese Texte unter der Rubrik ‚Karlsepik‘, die in diesem Setting als *Chanson de geste* mit der Heldenepik zusammen berechnet wurde, ohne hiermit Gattungsimplikationen tätigen zu wollen.

an anderer Stelle zeigt, wäre zu überlegen, ob hier eher ein stilistisches als ein gattungsspezifisches Merkmal vorliegt. Allerdings ist das gewählte Verfahren für eine solche Schlussfolgerung zu schematisch angelegt und das Vergleichskorpus (besonders im Bereich der Bibeldichtung) wohl auch nicht breit genug. Wichtig erscheint dies dennoch: Philipp wurde in der Forschung traditionell seine vermeintlich unbeholfene (Reim-)Sprache vorgehalten, was sicher einen Beitrag dazu geleistet hat, dass sich die zeitgenössische Wertschätzung des Textes bis heute nicht in der Beachtung seitens der germanistischen Forschung widerspiegelt.<sup>43</sup> Die hier herausgearbeiteten sprachlichen Wiederholungsstrukturen des ‚Marienlebens‘ sollen dies explizit nicht bestärken;<sup>44</sup> Philipp teilt seine wiederholende Erzählweise im Gegenteil mit einigen der wichtigsten Autoren der mittelhochdeutschen Literatur.<sup>45</sup>

## Literaturverzeichnis

### Quellen

Pommersfelden, Gräfl. Schönbornsche  
Schlossbibl., Cod. 46 (2797).

**Rückert, Heinrich (Hg.):** Bruder Philipps des  
Carthäusers Marienleben. Quedlinburg,  
Leipzig 1853.

### Forschungsliteratur

**Blumh, Lothar:** „zieml. correct u. fleißig“.  
Jacob und Wilhelm Grimm und das  
Marienleben des Kartäusers Bruder  
Philipp. Eine Adnote zur Wissenschafts-  
geschichte des frühen 19. Jahrhunderts.  
In: Zeitschrift für Germanistik 2 (1992),  
S. 371–375.

**Flecken-Büttner, Susanne:** Wiederholung  
und Variation als poetisches Prinzip.  
Exemplarität, Identität und Exzeptionali-  
tät in Gottfrieds ‚Tristan‘. Berlin 2011.

**Gärtner, Kurt:** Die Überlieferungsgeschichte  
von Bruder Philipps Marienleben.  
Marburg 1978.

<sup>43</sup> Vgl. Ostermann 2019, S. 85–87; vgl. auch Blumh 1992.

<sup>44</sup> Vgl. Lieb 2001, S. 507: „In anderer Hinsicht sind solche Zuweisungen problematisch, wenn sie nämlich (und implizit tun sie dies gerne) mit Bewertungen der Literarizität zusammenfallen: Die als rhetorisch, didaktisch, mündlich usw. ausgewiesenen Wiederholungen werden zu Merkmalen der Niederungen poetischer Produktion, gegen die sich die ‚eigentliche‘ Literatur abhebt.“

<sup>45</sup> Vgl. Lobsien 1995, S. 15: „Die Wiederholung ist das augenfälligste ‚technische‘ Kennzeichen poetischer Texte“; ferner Flecken-Büttner 2011, S. 3: „Immer wieder hat man Schönheit und Strukturierung oder Subversion von Ordnungen in den Kategorien Wiederholung und Variation oder, nimmt man die zeitliche Dimension zurück, Identität und Differenz beschrieben, hat man Dichtung durch eine besondere Dichte und Prägnanz entsprechender Relationen ausgezeichnet gesehen.“

- Gärtner, Kurt:** Zur Neuausgabe von Bruder Philipps Marienleben (ATB). In: Anton Schwob (Hg.): Editionsberichte zur mittelalterlichen deutschen Literatur. Beiträge der Bamberger Tagung Methoden und Probleme der Edition mittelalterlicher deutscher Texte. Göttingen 1994, S. 33–41.
- Gärtner, Kurt:** Marienverehrung und Marienepik im Deutschen Orden. In: Jaroslav Wenta, Sieglinde Hartmann u. Gisela Vollmann-Profe (Hgg.): Mittelalterliche Kultur und Literatur im Deutschordensstaat in Preußen. Leben und Nachleben. Torun 2008, S. 395–410.
- Gärtner, Kurt:** Prologversionen zu Philipps ‚Marienleben‘. In: Bernhart Jähning u. Arno Mentzel-Reuters (Hgg.): Neue Studien zur Literatur im Deutschen Orden. Stuttgart 2014, S. 137–146.
- Gay-Canton, Réjane:** Zwischen Zensur und Selbstzensur. Verbesserungappelle in der Vita beate Marie et Salvatoris Rhythmica und ihren mittelhochdeutschen Bearbeitungen. In: Barbara Fleith u. René Wetzl (Hgg.): Kulturtopographie des deutschsprachigen Südwestens im späteren Mittelalter. Studien und Texte. Berlin, New York 2009, S. 41–60.
- Hausmann, Albrecht:** Gott als Funktion erzählter Kontingenz. Zum Phänomen der ‚Wiederholung‘ in Hartmanns von Aue Gregorius. In: Cornelia Herberichs u. Susanne Reichlin (Hgg.): Kein Zufall. Konzeptionen von Kontingenz in der mittelalterlichen Literatur. Göttingen 2010, S. 79–109.
- Lieb, Ludger:** Eine Poetik der Wiederholung. Regeln und Funktionen der Minnerede. In: Ursula Peters (Hg.): Text und Kultur. Mittelalterliche Literatur 1150–1450. Stuttgart 2001, S. 506–528.
- Lobsien, Eckhard:** Wörtlichkeit und Wiederholung. Phänomenologie poetischer Sprache. München 1995.
- Masser, Achim:** Marien- und Leben-Jesu-Dichtung. In: Ingeborg Glier (Hg.): Die deutsche Literatur im späten Mittelalter 1250–1370. Zweiter Teil: Reimpaargedichte, Drama, Prosa. München 1987, S. 142–152.
- Ostermann, Christina:** Cum iubilationibus et dulci symphonia. Die Darstellung der Himmelfahrt Mariens in Bruder Philipps ‚Marienleben‘. In: Zeitschrift für deutsches Altertum und deutsche Literatur 148 (2019), S. 85–101.
- Reul, Christian, Stefan Tomasek, Florian Langhanki u. a.:** Open Source Handwritten Text Recognition on Medieval Manuscripts Using Mixed Models and Document-Specific Finetuning. In: 15th IAPR International Workshop on Document Analysis Systems (DAS). 2022, S. 414–428. [https://doi.org/10.1007/978-3-031-06555-2\\_28](https://doi.org/10.1007/978-3-031-06555-2_28) (Zugriff: 08.11.2024).
- Roßbach, Nikola u. Angela Schrott:** Wiederholung und Variation im Gespräch des Mittelalters und der Frühen Neuzeit. Einleitende Überlegungen. In: Dies. (Hgg.): Wiederholung und Variation im Gespräch des Mittelalters und der Frühen Neuzeit. Berlin, Boston 2023, S. 3–19.
- Schwarzbach-Dobson, Michael:** Wiederholung als Differenz. Paradoxien der Entzeitlichung im Reynke de Vos (1498). In: Kathrin Lukaschek, Michael Waltenberger u. Maximilian Wick (Hgg.): Die Zeit der sprachbegabten Tiere. Ordnung, Varianz und Geschichtlichkeit (in) der Tierepik (Beiträge zur mediävistischen Erzählforschung, Themenheft 11). Oldenburg 2022.

**Tomasek, Stefan, Christian Reul u.**

**Maximilian Wehner:** Handwritten  
Text Recognition und Word Mover’s  
Distance als Grundlagen der digitalen  
Edition „Die Kindheit Jesu Konrads

von Fußesbrunnen“. In: 8. Jahrestagung  
des Verbands Digital Humanities im  
deutschsprachigen Raum (DHd). 2022,  
S. 134–138. [https://doi.org/10.5281/  
zenodo.6304590](https://doi.org/10.5281/zenodo.6304590) (Zugriff: 08.11.2024).



## FORUM MITTELALTER

---





**Maryvonne Hagby**, ‚Die Königstochter von Frankreich‘ des Hans von Büchel. Fünf kontextualisierende Studien. Mit einer *Bibliographie raisonnée* zum Manekine-Stoff (Studien und Texte zum Mittelalter und zur frühen Neuzeit 23). Münster, Waxmann 2023. 577 S.

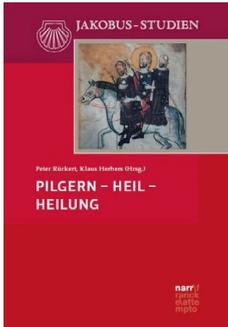
**Besprochen von Monika Unzeitig:**  
Greifswald, unzeitig@uni-greifswald.de

Der Versroman ‚Die Königstochter von Frankreich‘, im Jahr 1400 von Hans von Büchel verfasst, gehört zu den weniger bekannten Romanen zwischen Spätmittelalter und Früher Neuzeit und hat in der mediävistischen Forschung bislang kaum Beachtung gefunden. Handschriftlich ist der Roman nur fragmentarisch überliefert, aber seine Zweitrezption im frühen Druck durch Johann Grüninger (Straßburg 1500 und 1508) bezeugt ein andauerndes Interesse an dieser Erzählung. Maryvonne HAGBY hat 2017 eine Textausgabe nach der Druckfassung veröffentlicht und nun 2023 eine umfangreiche Monographie vorgelegt, die überhaupt die erste zur ‚Königstochter von Frankreich‘ ist. Der Text verbindet in eigenständiger Weise Fiktion und Historiographie, indem die fiktionale Handlung des Romans, und das ist die wesentliche Neuerung, mit einer zeitaktuellen historisierenden Schlusserzählung endet.

Mit Blick auf die Besonderheiten des hybriden Romans, der als Sonderfall in der Literaturgeschichte an der Schwelle zum 15. Jahrhundert, einer Zeit lückenhafter Romanproduktion, gelten darf, wählt die Verfasserin einen Zugang zu diesem Werk über seine literarische und historische Kontextualisierung. In fünf Einzelstudien, die sich thematisch und methodisch ergänzen, arbeitet sie heraus, wie der Autor einerseits mittelalterlichen Erzähltraditionen folgend den Erzähltypus von der unschuldig verfolgten und verleumdeten Jungfrau als Minne- und Aventureroman entwirft, andererseits durch gezielte Abweichungen und Neuerungen strategisch die Erzählung funktionalisiert, um so seiner politischen Zielsetzung entsprechend die Entstehung des Hundertjährigen Krieges durch eine gescheiterte Erbfolge zu erklären. Die umfassenden vergleichenden und kontextualisierenden Studien situieren den Roman zwischen Tradition und Transformation und explizieren durch die verschiedenen Zugriffe die Komplexität des Erzählkonzepts.

Der ‚Königstochter von Frankreich‘ ist keine direkte Vorlage zuzuordnen; inhaltlich gehört sie zum bekannten Manekine-Stoff (Studie I). Der Vergleich mit den Bearbeitungen des Stoffs (siehe auch die umfassende ‚Bibliographie raisonnée‘) zeigt, dass zwar Strukturtyp, Figurenkonstellation und konstitutive Motive aufgegriffen werden, jedoch das Erzählen vom Scheitern der genealogischen Nachfolge das Handlungsschema verlängert und keine (schemagerechte) glückliche Lösung bietet. Der kontextbezogene Blick auf das Genre des Minne- und Aventurierromans (Studie II) verweist auf eine deutliche Reduzierung gattungstypischer Erzählelemente und betont die singuläre Stellung des Romans durch die Verlagerung der Konflikte in den Bereich politischer Kommunikation. Die Liebe ist zudem nicht mehr handlungskonstitutiv, stattdessen zielt die deutliche Anlehnung an die hagiographische Literatur (Studie III) auf eine Figurenkonzeption, die die Königstochter durch ihre außerordentliche Frömmigkeit idealisiert, sie fast als Heilige inszeniert und so die Akzeptanz der Erzählung durch Rückgriff auf legendarische Erzählmodelle sichert. Damit rückt der Fokus auf die Rezeptionsteuerung: Die Autorisierung erfolgt nicht (mehr) durch traditionelle, auktoriale Mittel der Wahrheitsbeteuerung. Vielmehr soll durch die Erzählweise und die interaktive Einbeziehung der Rezipierenden (auch auf der emotionalen Ebene) Glaubwürdigkeit erzeugt werden, wie die Analyse der kommunikativen Akte zeigt, die durchgehend den Roman strukturieren (Studie IV). Die mündlichen und schriftlichen Formen kommunikativen Handelns zwischen den Figuren legen die Problemkonstellation offen, so dass wiederum durch die in der Textlektüre angelegte Wissensvermittlung eine plausible, verbürgte und zugleich persuasive Erzählung begründet werden kann. Explizit wird im Schlussteil der Krieg vereinfachend (und ahistorisch) als zwingende Folge der nicht gesicherten königlichen Nachfolge beschrieben – mit einer Parteinahme für die Engländer (Studie V). Dazu erfolgt die Vermittlung des historischen Wissens in einer doppelten Perspektivierung: im Duktus chronikalen Erzählens – dies dezidiert ohne Faktualitätssignale – und in der Tradition genealogischer Ursprungsgeschichten unter Rekurs auf eine mythologisierende Erzählweise; dies ganz in der Absicht, den Krieg mittels einer fiktionalen Geschichte zu erklären.

Die materialreichen Studien zu dem Versroman und seiner Sonderstellung dürften für künftige Untersuchungen Anregung und Fundgrube sein.



**Klaus Herbers u. Peter Rückert (Hgg.),** Pilgern – Heil – Heilung (Jakobus-Studien 25). Tübingen, Narr Francke Attempto 2023. 224 S., 14 Abb.

**Besprochen von Christian Kaiser:**  
Bonn, ckaiser@uni-bonn.de

Der Sammelband möchte eine Betrachtung des Themas ‚Pilgern zur Erlangung spirituellen Heils und körperlicher Heilung‘ aus unterschiedlichen Richtungen erreichen, die noch dazu eine Verbindung zwischen Mittelalter und aktuellen Bezügen herstellen sollen. Es sei dem Rezensenten gestattet, hierzu eine differenzierte Beurteilung abzugeben. Äußerst gelungen erscheinen die dezidiert historischen Beiträge, die sich allesamt auf das Mittelalter und die Frühe Neuzeit beziehen. Die liturgiehistorischen Beobachtungen von Jürgen BÄRSCH, der Aufsatz von Berndt HAMM über das spätmittelalterliche Pilgern im Kontext der anwachsenden Diesseits- und Jenseitsängste im Zeitalter nach der Großen Pest sowie Julia BURKHARDTS Text über das ‚geistige Pilgern‘ im Mittelalter sind sehr gut recherchierte, facettenreiche Studien mit reicher Quellendiskussion gemäß dem aktuellen Forschungsstand. Sie beleuchten eindrucksvoll die Perspektive derjenigen Personen, die sich Heil und Heilung versprochen, die individuellen und strukturellen Schwierigkeiten der Realisierung einer Pilgerreise sowie die gängigen Alternativen. Durchgehend nehmen sie Bezug auf die zeitgenössischen Vorstellungen und Erklärungen der erhofften Heilwirkungen, seien es seelisches Heil durch Buße, Umkehr und Versöhnung und der damit verbundene innere Wandlungsprozess mit einer neuen Ausrichtung auf Gott, seien es die Segensmacht der Heiligen, die wunderbaren Wirkungen auf Leib und Seele, die Verschonung von Krankheit und Unglück, die Gnadenmittel der Kirche, das nie ganz geklärte Verhältnis zwischen göttlicher Gnade und menschlicher Aktivität auf dem Heilsweg des Pilgers. Eingehend berücksichtigt werden auch immer die für das Pilger- und das Ablasswesen zuständigen Instanzen und deren interne Auseinandersetzungen, so dass hier Veränderung und Wandel gut nachvollziehbar werden. Auch

der Aufsatz ‚Heilung durch Ablass?‘ von Christiane LAUDAGE weiß in diesen Hinsichten zu überzeugen, wobei allerdings in ihrer überaus detailreichen Studie das Rahmenthema selbst kaum verhandelt wird, sondern quasi als nicht weiter erklärungsbedürftige Annahme, das Seelenheil sei im Spätmittelalter in erster Linie mit dem Ablass verbunden gewesen, vorausgesetzt wird. Insgesamt lässt sich sagen, dass die genannten vier geschichtswissenschaftlichen Beiträge künftig von allen Forschenden, die sich mit den Heilsaspekten des Pilgerns beschäftigen, sicherlich mit Gewinn gelesen werden.

Der Band enthält auch zwei Beiträge mit anderer thematischer Ausrichtung. Der Bericht von Beate BRIESECK über die ‚Jakobusgruppe‘, ein aktuelles therapeutisches Angebot einer psychiatrischen und psychotherapeutischen Klinik, und ihre Ausführungen zu den Effekten des Zurücklegens des Jakobuswegs in Teiletappen in einer Gruppe mit Menschen mit psychotherapeutischem Bedarf bieten sehr interessante Einblicke durch ihren praktischen Zugang. Zu den anderen Beiträgen des Sammelbands besteht allerdings kein Bezug, so dass sie in diesem Kontext isoliert erscheinen. Der kurze Text ‚Pilgern als Lebenskunst‘ des Philosophen Wilhelm SCHMID fällt im Vergleich zu den bisher genannten deutlich ab. Es handelt sich sowohl methodisch als auch inhaltlich eher um eine feuilletonistisch-persönliche Reflexion, die sich weder um Kontextualisierung der eigenen Hypothesen noch um eine Brücke zu anderen Beiträgen bemüht. Die beiden Herausgeber steuern neben einer sehr knappen Einleitung auch noch eigene Texte zur Rahmung der Sammlung bei: am Anfang Klaus HERBERS' Aufsatz ‚Heil oder Heilung. Wunder und Übernatürliches beim Pilgern‘ und am Ende Peter RÜCKERTS Darstellung von ‚Margarethe von Savoyen auf dem Weg nach Santiago de Compostela‘. Beide geben zum größten Teil bereits von ihnen selbst veröffentlichtes Material wieder. Insgesamt beschränkt der Sammelband seinen Zugang zum Phänomen Pilgern auf die Perspektive der Hoffenden (auf Heil und Heilung) und deren Regulierung bzw. Begleitung; nicht in Betracht gezogen werden etwa die Sichtweisen der nicht (mehr) Hoffenden oder der Enttäuschten. Durch diese Vor-Entscheidung fallen andere mögliche Blickrichtungen heraus, wie sie bspw. in Boccaccios ‚Decameron‘ in den Novellen II, 1 und III, 7 anklingen, nämlich dass im Rahmen der Heils- und Heilungserwartungen des (mittelalterlichen) Pilgers Heuchelei, Betrug und Camouflage durchaus vorkommen konnten (um es vorsichtig zu formulieren). Was ließe sich weiterhin aus den Quellen zu den sicherlich existenten Fällen erfahren, in denen (körperlich oder geistig) Pilgernde eben nicht ‚durch Wunder‘ gesund wurden? Ebenfalls nicht zur Sprache kommt, wie sich die zeitgenössischen Ärzte gegenüber dem Heilungsversprechen des Pilgerwesens verhielten, welche Chance bzw. welche Konkurrenz sie darin erkannten. In einem Band, der sich mit vor-modernen Konzepten von Heilung auseinandersetzt, wäre die Einbeziehung einschlägiger Stellungnahmen aus der Medizin der Zeit mindestens sinnvoll gewesen.



**Christoph Kann und Monika Steffens (Übers.),** Heinrich von Gent. Die Möglichkeit menschlichen Wissens, Artikel 1 der Summa Quaestionum Ordinariorum. Lateinisch-Deutsch. Übersetzt und eingeleitet von Christoph Kann und Monika Steffens (Herders Bibliothek der Philosophie des Mittelalters 59). Freiburg im Br., Basel, Wien, Herder 2024. 373 S.

**Besprochen von Julia Beier:** Bonn, jbeier1@uni-bonn.de

Bei dem vorliegenden Band aus ‚Herders Bibliothek der Philosophie des Mittelalters‘ handelt es sich um die von Christoph KANN und Monika STEFFENS verantwortete Übersetzung und kommentierende Einleitung des ersten Artikels der ‚Summa Quaestionum Ordinariorum‘ (im Folgenden ‚Summa‘) des Heinrich von Gent, eines Theologen und Philosophen der Hochscholastik (gest. 1293). Die Arbeit setzt mit einer Einführung zu Leben und Werk Heinrichs ein (7–46). Darin enthalten ist eine Analyse seiner großen Bedeutung für die Philosophiegeschichte. Seine Hauptwerke sind die ‚Quodlibeta‘, eine Zusammenstellung akademischer Disputationen in literarischer Form, sowie die ‚Summa‘, ein aus der Unterrichtspraxis hervorgegangenes Werk, dessen erster Artikel Gegenstand des hier besprochenen Buches ist (10–12). Heinrich kommt eine Vermittlerrolle zwischen dem in der Tradition verwurzelten Denken platonisch-augustinischer Prägung und einer stärker aristotelisch ausgerichteten Geisteshaltung zu: „Sein Rückgriff auf platonische ebenso wie auf aristotelische Lehrstücke [...] [ist] als Syntheseleistung zu würdigen [...]“ (17). Anhand der Betrachtung von Leben, Werk und philosophisch-theologischer Grundhaltung arbeiten die Autoren heraus, dass Heinrich damit einen völlig neuen Ansatz verfolgte, der für nachfolgende Generationen von Denkern eine wichtige Voraussetzung für die Weiterentwicklung philosophischer Ideen darstellte. KANN und STEFFENS zeigen, dass aufgrund der dominanten Rezeption der Werke z. B. Thomas’ von Aquin die moderne Rezeption der Werke Heinrichs lange Zeit seiner geistesgeschichtlichen Bedeutung nicht gerecht wurde. Erst in jüngerer Vergangenheit sei in der Forschung ein größeres Interesse an Heinrich zu verzeichnen (23).

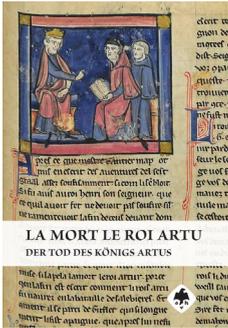
Im Rahmen der inhaltlichen Einführung wird ein Überblick über die Struktur des Textes und den Inhalt der zwölf ‚Quaestiones‘ gegeben (34–40). Es gelingt den

Autoren, zu zeigen, dass und wie die ‚Summa‘ von der philosophischen Problematik der Leistungsfähigkeit menschlicher Erkenntnis geprägt ist. Heinrich fragt grundsätzlich nach der Möglichkeit und Reichweite menschlichen Wissens vor dem Hintergrund göttlicher Einflussnahme und Begrenzung des Menschen und ist somit Wegbereiter eines späteren Skeptizismus (19). Die Einleitung schließt mit Bemerkungen zu Text und Übersetzung (40–46). Der abgedruckte lateinische Text basiert im Wesentlichen auf der kritischen Edition von G. A. WILSON aus dem Jahr 2005 (41).

Den Hauptteil des Buches (48–335) bildet die deutsche Übersetzung; ihr ist – den Gepflogenheiten der Reihe folgend – der lateinische Originaltext beigelegt. Der Apparat konzentriert sich auf die Quellen, in diesem Fall die Werke der Kirchenväter oder (pagan-)antiker und mittelalterlicher Autoren. Nur vereinzelt finden sich textkritische Anmerkungen (z. B. 70) oder inhaltliche Kommentierungen (z. B. 68), bisweilen auch Erläuterungen zur Übersetzung (z. B. 51). Änderungen am lateinischen Text nehmen KANN und STEFFENS lediglich im Falle offensichtlicher Druckfehler vor oder wenn sie in begründeten Einzelfällen einer anderen Lesart aus dem Apparat der Edition WILSONS folgen (41). Der Fokus liegt in dieser Arbeit darauf, eine vollständige lateinisch-deutsche Fassung des Anfangsartikels der ‚Summa‘ vorzulegen (40–46). Die Übersetzung orientiert sich an der Zielsprache und ist damit leicht verständlich. Trotz der seitens der Autoren angesprochenen Herausforderung durch die komplizierte Syntax Heinrichs ist der deutsche Text flüssig zu lesen (42).

Das Buch schließt mit einem umfangreichen Anhang (337–373), der die Abweichungen gegenüber WILSONS Edition anführt sowie textkritische Hinweise und eine Bibliographie umfasst. Hilfreich sind zudem ein Glossar, das die jeweilige Entscheidung der Autoren für die spezifische deutsche Übersetzung der wichtigsten von Heinrich genutzten lateinischen Vokabeln enthält, und ein Personen- sowie Stellenregister.

Insgesamt handelt es sich um eine Übersetzung mit Kommentar, die für eine große Bandbreite von Rezipienten aus Philosophie, Theologie und Geschichtswissenschaft gewinnbringend ist, da sie die Bedeutung von Heinrichs Leben und Werk verdeutlicht und sicherlich dazu beitragen wird, seinen Platz in der Philosophiegeschichte dauerhaft zu sichern. Es wäre wünschenswert, wenn diese gelungene Arbeit eine Fortsetzung finden würde.



**Fritz Peter Knapp (Übers.),** *La mort le roi Artu. Der Tod des König Artus.* Text nach der Ausgabe von Jean Frappier 1964. Übersetzung mit Anmerkungen und Einleitung von Fritz Peter Knapp. Stuttgart, Hiersemann 2023. XLV, 421 S.

**Besprochen von Marion Darilek:**

Tübingen, marion.darilek@uni-tuebingen.de

Fritz Peter KNAPPS eingeleitete und kommentierte Übersetzung von ‚La mort le roi Artu‘ (‚Der Tod des König Artus‘) erschließt den dritten und letzten Teil des altfranzösischen Prosaromanzyklus von ‚Lancelot und dem Graal‘ (1. Drittel des 13. Jahrhunderts) erstmals philologisch genau einer deutschsprachigen Leserschaft. Während mehrere englische und neufranzösische Übersetzungen von ‚La mort le roi Artu‘ vorliegen, existierte im Neuhochdeutschen neben der zweisprachigen Edition des mittelhochdeutschen ‚Prosalancelot‘ (‚Bibliothek deutscher Klassiker‘, 1995–2004) bislang nur eine freie Nacherzählung (Ruth SCHIRMER, 1961). Dies erstaunt, da es sich bei dem Roman um eine Hauptquelle für Thomas Malorys mittelenglischen Prosaroman ‚Le Morte Darthur‘ (15. Jahrhundert) handelt, der das Artusbild vor allem der angelsächsischen Tradition bis heute prägt (VII–X). Entsprechend beklagt KNAPP, dass „sich [...] in der deutschen Literatur vom Hochmittelalter bis heute das famose Ende der Geschichte von König Artus und seiner Tafelrunde [...] nur in schattenhaften Umrissen ab[zeichnet]“, und verbindet mit seiner Übersetzung implizit den Anspruch, die „Bildkraft“ des „schicksalhafte[n] Untergang[s]“ (X) ins Deutsche zu transportieren.

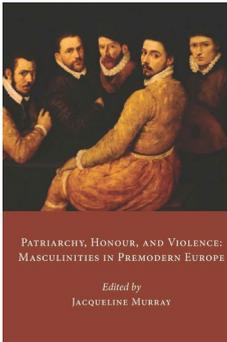
Die Einleitung verortet ‚La mort le roi Artu‘ zunächst vor dem weiteren Hintergrund der arthurischen Tradition (VII–X) sowie – verbunden mit Zusammenfassungen des ‚Lancelot A‘ und ‚B‘ und der ‚Queste del Saint Graal‘ – im engeren literarhistorischen Kontext des ‚Lancelot-Graal-Zyklus‘ (X–XXIII). Hier schon wirft KNAPP zentrale Fragen auf, indem er u. a. politische Kontexte, das Liebeskonzept, den religiösen Gehalt, die (fingierte) Autorschaft sowie die Historizität (so die mittelalterliche Annahme) ‚König Artus‘ und des Romanzyklus beleuchtet. Diese Aspekte werden in Bezug auf „Gattung und Gehalt“ (XXVI–XXXVI) vertieft, wobei

KNAPP für den Gesamtzyklus sowohl „kleinere Widersprüche im Detail“ als auch „umfassende Klammern“ (XXXVI) herausarbeitet. Konzise dargestellt sind „Aufbau und Stoff, Sprache und Stil“ (XXIII–XXVI) von ‚La mort le roi Artu‘.

Als Textgrundlage wählt KNAPP die Ausgabe von Jean FRAPPIER (<sup>3</sup>1964), deren Basis die Leithandschrift A aus dem ersten Handschriftenzweig ist. Die Entscheidung gegen neuere Editionen von Handschriften des zweiten Zweigs begründet er einerseits überlieferungsgeschichtlich, andererseits urheberrechtlich (XXXVII f.). Bei der Untergliederung des Romans übernimmt er erstens die Paragraphen von FRAPPIER, die meist den Initialen in Handschrift A folgen (XXXVIII). Diese fasst er zweitens zu übergreifenden „Erzählpartien“ zusammen, die sich an „vom Autor stammenden Signale[n]“, die einen „Wechsel der Akteure“ (XXIII) anzeigen, orientieren. Zudem ergänzt KNAPP drittens eine fünfteilige Kapiteleinteilung (XXIV). Trotz dieser plausiblen Gliederung fällt die Orientierung im 421 Seiten umfassenden Text mit Übersetzung jedoch schwer, da lediglich unscheinbare Zwischenüberschriften die Struktur verdeutlichen. Weder untergliedert das Inhaltsverzeichnis den Text- und Übersetzungsteil weiter noch werden lebende Kolummentitel verwendet.

Seine Übersetzungsweise macht KNAPP transparent, indem er Vorentscheidungen wie z. B. Tempus-Wechsel, Inquit-Formeln, wiederkehrende Vokabeln schon einleitend offenlegt (XXXVIII–XLII), aber auch an Einzelstellen semantische Eigenheiten des Altfranzösischen und schwierige Passagen kommentiert und mit anderen Übersetzungen abgleicht. Hinzu kommen Sachkommentare u. a. zu Figuren und Orten aus den anderen Teilen des ‚Lancelot-Gral-Zyklus‘. Die Übersetzung ist präzise, der Sprachstil klassisch-gehoben. Die teils verschachtelten Sätze sind der altfranzösischen Syntax geschuldet. Mitunter hätte eine etwas zeitgemäßere Wortwahl dazu beitragen können, die ‚Bildkraft‘ der Erzählung einem heutigen Publikum besser zu vermitteln.

Das Verdienst von KNAPPS Übersetzung, die schon durch ihre Länge und die souveräne Navigation durch die komplexe Handlung des ‚Lancelot-Gral-Zyklus‘ beeindruckt, liegt darin, dass sie das Repertoire komparatistischer Untersuchungen zum arthurischen Roman um einen altfranzösischen Text erweitert, dem eine Schlüsselrolle für die Mittelalterrezeption zukommt. Denn dem zunehmenden mediävistischen Interesse an sprachvergleichenden Studien steht nicht selten die Hürde fehlender Sprachkenntnisse entgegen. Substantiell eingeleitete und kommentierte zweisprachige Textausgaben sind daher unerlässlich, um die französischsprachige Literatur des Mittelalters auch nicht romanistisch geschulten Mediävistinnen und Mediävisten sowie Studierenden zugänglich zu machen.



**Jacqueline Murray (Hg.)**, *Patriarchy, Honour, and Violence. Masculinities in Premodern Europe* (Essays and Studies 57). Toronto, Centre for Renaissance and Reformation Studies 2022. 358 S. 21 Abb.

**Besprochen von Simon Suttmann:**

Freiburg, [simon.suttmann@grk2571.uni-freiburg.de](mailto:simon.suttmann@grk2571.uni-freiburg.de)

Der Band ist aus einer im November 2020 in Toronto unter dem Titel ‚Masculinities in the Premodern World: Continuities, Change, and Contradictions‘ abgehaltenen Konferenz hervorgegangen und vereint 13 Aufsätze, die unterschiedliche Aspekte des Phänomens ‚Maskulinität‘ anhand von Fallstudien herausarbeiten. Diese haben dabei jeweils einen zeitlich, räumlich und thematisch relativ eng gewählten Fokus und ziehen zumeist ein klar definiertes Quellenkorpus zur Analyse heran. Zusammengenommen sind zwei regionale Schwerpunkte erkennbar: Nord- und Mittelitalien (sieben Beiträge) und England (drei Beiträge). Daneben finden sich Ausflüge nach Flandern und auf die Iberische Halbinsel. Zeitlich wird eine Spanne vom 11. bis zum 16. Jahrhundert abgedeckt, wobei das späte Mittelalter dominiert.

Inhaltlich gliedert sich der Band in drei Teile. Der erste Teil ist dem Thema ‚Tensions in Hegemonic Masculinity‘ gewidmet. Ihm liegt die Überlegung zugrunde, dass Ehr- und Männlichkeitskonzepte eng miteinander verknüpft waren und in Form von Dominanzverhalten performativ zur Geltung gebracht werden mussten. Die entsprechenden Beiträge analysieren daher in erster Linie Konflikte, die entstanden oder zur Eskalation getrieben wurden, weil spezifische Vorstellungen von männlicher Ehre das Verhalten der beteiligten Personen und Personengruppen (mit)bestimmten. Hervorzuheben ist in diesem Zusammenhang etwa der Aufsatz von Aaron MIEDEMA, der anhand von Gerichtsdokumenten minutiös nachzeichnet, wie im Latium des späten 16. Jahrhunderts ein Rechtsstreit zwischen einem Geistlichen und einem Laien in eine gewaltsame Auseinandersetzung mündete, weil beide Parteien wechselseitig Angriffe auf die männliche Ehre ihres Gegenübers unternahmen. Shannon McSHEFFREYS Versuch, Feindschaften zwischen italienischstämmigen und englischen Geschäftsleuten im London des

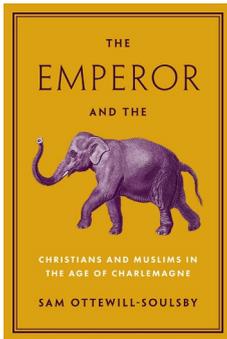
ausgehenden Mittelalters auf unterschiedliche Vorstellungen von männlicher Sexualmoral zurückzuführen, zeigt zudem auf, dass sich Spannungen gerade auch dann ergeben konnten, wenn konkurrierende Deutungsmuster für männliches Dominanzverhalten zur Verfügung standen.

Der zweite Teil unter dem Titel ‚Men in Groups: Homosociality‘ beschäftigt sich mit Nahbeziehungen zwischen Männern und verschiedenen Formen spezifisch männlicher Gruppenbildung bzw. Vergemeinschaftung. Obwohl keine wirklich überzeugende Eingrenzung des Themenbereichs erfolgt, gelingt es den einzelnen Autor\*innen doch, eine große Bandbreite sozialer Praktiken zu beleuchten und vor dem Hintergrund vormoderner Männlichkeitsvorstellungen zu deuten. Vor allem der Beitrag von Savannah PINE zur Ritualbruderschaft unter spätmittelalterlichen Rittern nimmt den Anspruch ernst, spezifische Formen der ‚Homosozialität‘ zu untersuchen. Der für sich betrachtet sehr interessante und überzeugende Beitrag von Andrew LARSEN zu Krisen männlicher Rollenbilder als Ursache für deviantes Verhalten bei mittelalterlichen Studenten der Universität Oxford hätte allerdings ebenso gut unter der Rubrik des vorherigen Abschnitts geführt werden können.

Der letzte Teil des Buches trägt den Titel ‚Patriarchy, Honour, and Lineage‘. Die dort versammelten Aufsätze eint die Frage, welche Rolle die Führung eines ‚Haushalts‘ und das Hervorbringen einer biologischen oder sozial konstruierten Nachkommenschaft bei der Konstituierung von Männlichkeit spielten.

Insgesamt findet die/der Leser\*in in dem Band eine Reihe anregender Fallstudien vor, die unterschiedliche Fragestellungen zum Thema Männlichkeit an teils sehr interessanten (und mitunter unterhaltsamen) Fallbeispielen durchexerzieren. Vor diesem Hintergrund ist es auch von Vorteil, dass der Band offenbar bewusst nicht auf ein bestimmtes Konzept von Männlichkeit festgelegt ist und weitgehend offenlässt, was genau unter Begriffen wie ‚masculinities‘, ‚patriarchy‘ oder ‚honour‘ zu verstehen sei und was die Autor\*innen damit zu assoziieren haben. Ein Nachteil dieser Herangehensweise ist jedoch, dass dadurch eine Zusammenführung der Begriffe und eine übergreifende Diskussion über „continuities, change, and contradictions“ (7) schwerfällt. Aus diesem Grunde wäre es für Leser\*innen, die sich für die Konjunkturen vormoderner Männlichkeitsvorstellungen interessieren, sicherlich hilfreich gewesen, eine ausführlichere theoretische Einführung oder ein Schlusskapitel hinzuzufügen, in dem die Ergebnisse der Konferenz bzw. des Bandes in der Gesamtschau diskutiert werden.

Als Ergänzung zu dem behandelten Buch dürfen sich solche Leser\*innen dafür über zwei weitere Bände freuen, die die Herausgeberin in jüngster Zeit zum Thema Männlichkeit in der Vormoderne veröffentlicht hat: ‚The Male Body and Social Masculinity in Premodern Europe‘ (2022) und zusammen mit Konrad EISENBICHLER ‚Premodern Masculinities in Transition‘ (2024).



**Sam Ottewill-Soulsby**, *The Emperor and the Elephant: Christians and Muslims in the Age of Charlemagne*. Princeton, Oxford, Princeton University Press 2023. 392 S. 9 Abb.

**Besprochen von Eric Böhme:**  
Dresden, eric.boehme@tu-dresden.de

Abul Abaz († 810) kann wohl als berühmtester Elefant des europäischen Mittelalters gelten. Für die vorliegende Druckfassung einer 2017 verteidigten Dissertation ist das 801 als diplomatisches Geschenk des ‘Abbāsiden-Kalifen Hārūn al-Rašīd zu Karl dem Großen geschickte Tier aber eher ein plakatives Emblem für eine Zusammenschau der diplomatischen Beziehungen zwischen den Karolingern und der arabisch-islamischen Sphäre. Sam OTTEWILL-SOULSBYS Beitrag zu diesem seit langem diskutierten Forschungsfeld konzentriert sich nicht nur auf die Beziehungen zwischen den Herrscherhöfen, sondern nimmt auch die Vielzahl nachgeordneter Akteure in den Blick, die an den Peripherien der imperialen Räume ihre eigene Agenda verfolgten. Damit sollen zum einen muslimische Perspektiven stärker in den Vordergrund gerückt und zum anderen mit der von F. W. BUCKLER bereits 1931 postulierten und schon damals umstrittenen These eines „Allianzsystems der vier Großmächte“ – Karolinger und ‘Abbāsiden gegen Byzantiner und Umayyaden – aufgeräumt werden (vgl. 6–8).

Nach der ausführlichen Einleitung behandelt Kapitel 2 die ideologischen und politischen Rahmenbedingungen karolingisch-muslimischer Diplomatie, die in den Kontext der politischen, soziokulturellen und ökonomischen Lage im Euromediterraneum eingeordnet werden. In Kapitel 3 geht es dann um die karolingischen Beziehungen zum ‘Abbāsiden-Kalifat. Hier wird schlüssig argumentiert, dass weder BUCKLERS vermeintliches Allianzsystem noch die *Terra Sancta* im Mittelpunkt der Kontakte standen. Vielmehr ging es beiden Seiten vorrangig um „Prestige-Diplomatie“ (30) in der Nachfolge Roms und Persiens, die die jeweiligen Herrscher zum Ausbau ihrer innenpolitischen Autorität nutzen konnten.

Die nachfolgend behandelten Beziehungen der Karolinger zu den andalusischen Umayyaden werden in zwei chronologisch geordnete Kapitel aufgeteilt.

Anders als im vorhergehenden Kapitel kann der Autor das Rad hier nicht neu erfinden, schafft es aber, aus dem spärlichen und häufig ambivalent zu deutenden Quellenmaterial eine kohärente und mit Blick auf die Akteure und deren Handlungslogiken deutlich nuanciertere Analyse der mehr als ein Jahrhundert andauernden militärischen und diplomatischen Kontakte zu formen. Zwischen der karolingischen *Marca Hispanica* und der umayyadischen *Tağr al-A'lā* habe, so das Urteil des Autors, eine transpyrenäische „frontier diplomacy“ (30) stattgefunden, in deren Rahmen Friedensschlüsse nie dauerhaft, sondern stets nur Zwischenlösungen gewesen seien.

Kapitel 6 widmet sich dann dem islamisch beherrschten, zentralen Mittelmeerraum zwischen Süditalien und Nordafrika als geographischer Leerstelle. Kontakte der Karolinger zu muslimischen Akteuren lassen sich hier nur sehr sporadisch belegen, fanden vorrangig in Zeiten exzeptioneller Krisen statt und wurden nie über längere Zeit aufrechterhalten. Die Ursache sieht der Autor darin, dass die Karolinger dort schlicht keine ebenbürtigen Gesprächspartner gefunden hätten. So habe man selbst vergleichsweise nahegelegene Herrschaftskomplexe wie das Emirat von Bari nicht als ernsthafte Bedrohung wahrgenommen und sie auch nicht als Verbündete gegen etwaige Rivalen auf der Halbinsel einsetzen können. Ludwig II. habe das Emirat 871 schließlich angegriffen und erobert, um durch die Vernichtung des vermeintlichen religiösen Gegners seine eigene Machtposition zu stärken.

In der kurzen Zusammenschau plädiert der Autor mit Recht dafür, die karolingisch-muslimischen Beziehungen nicht auf die Person Karls des Großen zu reduzieren, seien sie nach dessen Tod doch mit wechselnder Intensität fortgesetzt worden, woran beide Seiten ihren Anteil hatten. Das Karolingerreich dürfe nicht als rein europäische, sondern müsse viel stärker auch als mediterrane Macht begriffen werden, wofür ein geographisch weitwinkliger Blick erforderlich sei. Diesem Anspruch wird der Autor in der Tat gerecht, obgleich fraglich ist, ob die behandelten diplomatischen Kontakte in den Außenbeziehungen aller Beteiligten wirklich den hohen Stellenwert hatten, den er ihnen beimessen möchte. Durch eine vergleichende Betrachtung der muslimisch-byzantinischen Kontakte – die bekanntlich mit ganz anderer Intensität und Routine stattfanden – hätten die hervorragenden Analyseergebnisse etwas mehr kontextualisiert und letztendlich nivelliert werden können. Dennoch bietet OTTEWILL-SOULSBYS Studie nicht nur eine detaillierte Zusammenschau des nach wie vor intensiv diskutierten Themenfelds, sondern fügt der Forschungsdiskussion an zahlreichen Stellen nuanciertere oder gänzlich neue Impulse bei.



**Hans-Joachim Schmidt**, *Neue Gesetze für ein besseres Leben? Überlegungen zu Praktiken und Theorien der Gesetzgebung* (Scrinium Friburgense 58). Wiesbaden, Reichert 2024. 303 S.

**Besprochen von Felix Grollmann:**  
München, felix.grollmann@lmu.de

Große Abhandlungen über mittelalterliche Denkweisen vom Recht, wie sie Fritz KERN (1916/19), Ernst KANTOROWICZ (1957), Sten GAGNÉR (1960) oder Harold J. BERMAN (1983/2003) vorgelegt haben, sind rar. Ein Buch, das Theorien und Praktiken über die gesamte Epoche des westeuropäischen Mittelalters hinweg abdeckt, macht daher neugierig.

Der Vf. will wissen, ob es im Mittelalter nicht nur faktisch technologische und kulturelle Entwicklungen gab, sondern wie man sich ein gutes Leben im Diesseits vorstellte; ob sich Menschen zutrauten, wirksam Veränderungen hin zu einem besseren Leben durch Recht anstoßen zu können und ob sogar eine Idee von ‚Fortschritt‘ *avant la lettre* zu greifen ist. Trotz der originellen grundsätzlichen Fragestellung erliegt die Arbeit nicht der Versuchung, ein neues Meisternarrativ zu präsentieren – denn es soll „eine Vielstimmigkeit zum Klingen gebracht werden, die dissonante Töne hervorbringt“ (35).

Als wichtigste Einflüsse auf mittelalterliche Rechtskulturen werden in der insgesamt tendenziell chronologisch geordneten Darstellung das Christentum (37–62) sowie „Polis, Staat und Gesetz in der Antike“ (63–93) vorgestellt. Im anregenden Kapitel über die „Entstehung und Bewahrung von Recht im Mittelalter“ (95–127) – das eigentlich nur die erste Hälfte der Epoche abdeckt – streift der Vf. viele Quellengattungen sowie Leitbilder, z. B. *iusdictio* als königliche Kompetenz oder *necessitas temporum* als legitimierender Topos für Normänderung (112–116). Auch schwierige Quellen wie die frühmittelalterlichen ‚Leges‘ werden nicht monotonisiert, sondern deren Differenzen plastisch benannt. Methodisch eng werden nahezu ausschließlich Prologe ausgewertet; hier hätte der Regelungsgehalt der einzelnen ‚Leges‘ einbezogen werden können, der tiefere Einsichten ermöglicht

(z. B. ‚Lex Baiuvariorum‘ XIX, 8). Für die Frage diesseitiger Lebensoptimierung liefert das Material jedenfalls nicht viel Überraschendes. Hier führt das Kapitel, das theologische und geschichtliche Erwartungshaltungen im ausgehenden Hochmittelalter schildert, weiter (129–141). Im 12. Jahrhundert seien Vorstellungen aufgekommen, „dass das Heil bereits im diesseitigen menschlichen Leben eintreten würde“ (131).

Es folgen Ausführungen zur „Wiederentdeckung der Gesetzgebungen seit dem 12. Jahrhundert“ (143–205). Anfänge eines neuen Rechtsdenkens sieht der Vf. in der Zeit der Ausbildung des päpstlichen Primats. Symptomatisch für die umsichtige Abwägung ist die Beurteilung des Verhältnisses der Rechtswissenschaft zur Theologie als zwiespältig, da sowohl stärkere Unwandelbarkeit als auch Flexibilisierung zu konstatieren seien (158). Nicht ganz überzeugend ist eine Parallelisierung von ‚Liber Augustalis‘ (1231) und ‚Maiestas Carolina‘ (1355), wogegen das strukturverwandte Oberbayerische Landrecht Kaiser Ludwigs IV. (1346) als Exempel fürstlicher Gesetzgebung provinzialisiert wird. Disruptiver als herrscherliche Gesetzgebung habe das genossenschaftliche Recht der italienischen Stadtstaaten gewirkt, das seinen innovativen Charakter teils hinter Privilegien und Behauptungen von Kontinuität verbarg. Mit den Rechtsbüchern beschreibt der Vf. eine weitere klandestine Form der Bildung neuen Rechts, nur dass diesmal Experten nicht Rat gaben, sondern selbst die Feder führten, was eine Fortführung verunmöglicht habe.

In den „Theorien zum besseren Gesetz“ (207–272) zeigt der Vf. u. a., wie weltliche Juristen über neue und bessere Gesetze diskutierten. Die antike Herkunft des Römischen Rechts beförderte dabei ein Nachdenken über die Anpassung von Normen. Ungewöhnlich sind die Zitate des Vf. aus der Legistik mit bloßer Folio-Angabe alter Drucke, anstelle, wie üblich, die Referenzstelle im Römischen Recht mitaufzuführen. Der Vf. macht ein breites Spektrum sichtbar: Jacques de Révigny schließe Gesetze zum Ziel der Lebensbesserung aus, während Baldus de Ubaldis die kaiserliche Gesetzgebungsbefugnis sogar als „Motor für die Steigerung des allgemeinen Wohls“ (214) eingesetzt habe. Unter den Philosophen fällt Nicolas d’Oresme auf, der in getrennten Kapiteln Argumente für bzw. gegen Gesetzesänderungen gebündelt habe – die Entscheidung über die richtige Position überließ er dem König als Adressat seines Texts.

Wenige Kritikpunkte lassen sich vorbringen. Insgesamt liegt eine beeindruckende, schön geschriebene Abhandlung vor, die Normativität in zahllosen Facetten vorführt. Das Buch führt den Leser zu vielen ergiebigen Quellen hin. Überzeugend ist das Ergebnis, dass Rechtsveränderung häufig vorkam und nicht (immer) versteckt wurde. Die noch wichtigere Leistung des Werks liegt darin, zu zeigen, dass die Besserung des diesseitigen Lebens durch Rechtsetzung zumindest für einige Denker und Gesetzgeber eine ernsthafte Option war. Inwiefern dies auch im jeweiligen Regelungsgehalt konkret greifbar war, müsste aber noch untersucht werden.



**Jörg Sonntag**, *Der spielende Gott des Mittelalters. Eine kulturhistorische Annäherung (Mittelalter-Forschungen 70)*. Ostfildern, Thorbecke 2024. 442 S. 51 Abb.

**Besprochen von Martin Thurner:**  
München, thurner@lmu.de

Das Mittelalter als „Kulturgeschichte des spielenden Gottes“ (191)? Dass über das Phänomen des Spieles das ganze (abendländische) Mittelalter immanent erschlossen und in seiner Bedeutung für die weitere Kulturentwicklung im Allgemeinen und die Spieltheorie im Besonderen gewertet werden könnte, gehört nicht zu den gängigen Vorstellungen von jener Epoche, auch nicht innerhalb der wissenschaftlichen Mediävistik. Nach Lektüre des vorliegenden Bandes sollte man sagen, wünschen und hoffen: *noch* nicht! In seinem magistralen Opus gelingt Jörg Sonntag mit geradezu spielerisch virtuosem Ernst der Nachweis, dass im Mittelalter nicht nur auf ausnahmslos allen Feldern des religiösen, kulturellen und sozialen Lebens unbändig gespielt wurde, sondern dass dieser Spielrausch mit der Idee vom *deus ludens* auch im Mittelpunkt des mittelalterlichen „système divin“ (23) theoretisch wie lebenspraktisch begründet und mimetisch kulturstiftend wurde, auf eine Weise, in der das Mittelalter einen ebenso originellen wie bedeutsamen Beitrag zur allgemeinen Spieltheorie hinterlassen konnte.

Dieser gewichtige Ertrag zu dem (nur) an sich leichten Sujet gelingt dem Verfasser, indem er eine überwältigende Menge an empirischen Befunden (samt 51 Abbildungen) mit einer bestechenden Kraft zur systematischen Durchdringung und Synthese bündelt, kanalisiert und relativierend einordnet, wobei stets berücksichtigt bleibt, dass auch systematische Ideen ihre wechselvolle Geschichte haben und nicht fertig vom Himmel fallen, selbst oder gerade wenn das Spiel im Spiel ist.

Um den Gegenstandsbereich zugleich einzugrenzen und zu eröffnen, verbindet und kontrastiert der Autor die mittelalterlichen Sprachbefunde mit den Analysen der modernen Spielbegriffe, mit dem überraschenden Ergebnis, dass Wittgensteins Einsicht in die undefinierbarkeit von allgemeinen Eigenschaften

aller (Sprach-)Spiele durchaus der mittelalterlichen Situation nahekommmt. Im Mittelalter wird die Spielvielfalt aber nicht primär durch Analogien rund um einen gemeinsamen Bezugspunkt oder alternative ‚Familienähnlichkeiten‘ zusammengehalten, sondern durch ein dynamisch sich potenzierendes Ineinanderspiel von Faktizität und Allegorie.

Sowohl historisch wie auch systematisch kann man dies an der biblisch-patristischen Vorstellung vom in und mit der Schöpfung spielenden Gott theologisch aufhängen, an der heilgeschichtlichen Epiphanie im Gott-menschlichen Kind Jesus Christus anthropologisch ausbuchstabieren, in konkreten Spielen „präsenzsymbolisch“ (286) erproben, um alles dann wieder allegorisch auf den „(rück-)bespielten“ (321) Gott zurückzuprovozieren, und dies von der mystischen Innerlichkeit bis zu den städtischen Spelunken, sowohl in systemstärkender wie in systemsprengender Absicht. Durch die positive Integration des allgemein-ethischen Konzepts der *eutrapelia* (heitere Ausgewogenheit) aus der aristotelischen Ethik in die ‚Theologische Summe‘ etwa eines Thomas von Aquin kann der mittelalterliche *deus ludens* auch außerhalb sakraler Räume kulturprägend werden, was natürlich vor allem im Spätmittelalter mit einer Moralisierung und einer Unterscheidung von guten und schlechten, heiligen und dämonischen Spielen einhergehen muss – die Kontrastfolie traditioneller christlicher Spielkritik bleibt so produktiv erhalten. Kein Wunder, dass in der Mitte des Bandes dann das 1432 abgeschlossene ‚Guldin Spil‘ des Dominikaners Ingold von Basel durchaus überzeugend als eine Art ‚Summa ludica‘ einer ganzen Epoche präsentiert und interpretiert werden kann (235–289).

Die Erschließungskraft dieser integrativen Spiel-Optik für das Mittelalter liegt nicht zuletzt darin, dass es so gelingt, Ideen- mit Kultur- und auch Sozialgeschichte mühelos zu verbinden und damit die Vereinseitigung auf eine rein theologisch-dogmatische oder nur historisch-faktische Betrachtung dieser Epoche zu vermeiden. Spielen wurde – im mystischen Sinn bis ganz „dingallegorisch“ (327) oder gar als „in Epiphanien geschautes faktisches Tun Christi“ (103) – zur „Technik der Verfügbarmachung des Unverfügbaren“ (232) und zum „Seelentraining“ (37). Doch es erschöpfte sich nicht in dieser theologischen Sinngebung, denn „im Vergleich zu anderen Gottesfacetten“ ist das Motiv vom *deus ludens* „nicht nur immens wandlungs- und anpassungsfähig, sondern je nach Sachlage bausteinartig mit weiteren Trägern kultureller Codes kombinierbar“ (372). Weil sowohl Individuen (besonders Frauen) als auch Institutionen (Klöster, Höfe, Stadtgesellschaften) sich praktisch nachahmend vom *deus ludens* her ausleben konnten, wurde eine theologische Idee so zu einer „kulturellen Matrix“ (163), die damals ganze Welten durchlässig vereinte und auseinander hervorbrachte – und im Rückblick offenbar nur darauf wartete, heute erstmals umfassend interdisziplinär rekonstruiert zu werden.

Das Mittelalter  
Perspektiven mediävistischer Forschung  
2025 · Band 30 · Heft 1

Herausgegeben von Isabelle Mandrella  
im Auftrag des Präsidiums  
des Mediävistikverbandes

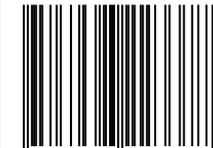


Das Themenheft widmet sich der computationell arbeitenden Mittelalterforschung in ihrer verbindungsstiftenden Funktion. Als Schnittstelle vermittelt sie zwischen den Disziplinen, Digitalem und Analogem, heterogenen Forschungsgegenständen und vielfältigen Methoden. Dabei integriert sie nicht nur Ansätze und Wissen aus anderen Fachbereichen, sondern trägt auch zum Erkenntnisgewinn über ihre eigene Disziplin hinaus bei. Das Ziel des Bandes ist, kollaborative Verfahren der *Digital Humanities* forschungsorientiert und praxisnah vorzustellen und die hierfür angewandten Methoden und Praktiken zu evaluieren und zu diskutieren.



**UNIVERSITÄT  
HEIDELBERG**  
ZUKUNFT  
SEIT 1386

ISBN 978-3-96822-337-7



9 783968 223377