

In Zukunft mit KI

Aus Anlass des 25-jährigen
Jubiläums des
Standardisierungsausschusses
Dezember 2025

Die Dokumentation

Das STA-Kolloquium wurde veranstaltet und unterstützt von

Bibliotheksservice-Zentrum Baden-Württemberg (BSZ)

Deutsche Nationalbibliothek (DNB)

Hessisches Bibliotheksinformationssystem (hebis)

Hochschulbibliothekszentrum NRW (hbz)

Schweizerische Nationalbibliothek

Staatsbibliothek zu Berlin (SBB-SPK)

Verbundzentrale des Gemeinsamen Bibliotheksverbundes (GBV VZG)

„Das Kolloquium soll motivieren, mutig und offen neue Schritte zu gehen - für eine Zukunft, in der KI und bibliothekarische Standards Hand in Hand gehen. “

Dr. Silke Schomburg (hbz) und Vorsitzende des Standardisierungsausschusses

„KI kann Erschließung, Metadatenpflege und Normdatenarbeit grundlegend verändern - entscheidend wird sein, wie wir technologische Möglichkeiten mit gemeinsamen Standards, Verantwortung und institutioneller Zusammenarbeit verbinden.“

Frank Scholze, Generaldirektor der DNB und Mitglied des Standardisierungsausschusses

Aus den Grußworten der Gastgeber

In Zukunft mit KI

ein Überblick

Aus Anlass des des 25-jährigen Jubiläums des Standardisierungsausschusses fand am 10. Dezember 2025 das

STA-Kolloquium "In Zukunft mit KI"

an der Deutschen Nationalbibliothek, Frankfurt, statt. Aus dem Kreis der Mitgliedsorganisationen wurden unterschiedliche Ansätze für den zukunftsorientierten Einsatz von Künstlicher Intelligenz (KI) in der Erschließung und Qualitätssicherung vorgestellt. Vor diesem Hintergrund diskutierten die Teilnehmenden des Kolloquiums die strategischen Implikationen eines möglichen Kulturwandels für die Bereitstellung und Vernetzung von Norm- und Metadaten durch KI.

Svein Arne Brygfjeld von der Norwegischen Nationalbibliothek stimmte die Teilnehmenden mit einer pointierten Keynote auf die strategische Dimension des Einsatzes von KI im bibliothekarischen Informationsmanagement ein.

Der fachliche Teil des Kolloquiums legte den Fokus auf KI im Kontext von Norm- und Metadaten. Es wurden verschiedene Einblicke in die Anwendung von KI-Werkzeugen als Unterstützung der Inhaltserschließung in Bibliotheken und verwandten Sparten wie Museen vorgestellt. Darüber hinaus berichtete man über Ansätze der Integration von KI-gestützten Ergebnissen in die vorhandenen Workflows. Schließlich wurden Beispiele für den Einsatz von KI-Technologien in Discovery-Systemen und die Rechercheberatung präsentiert.

Der zweite Teil des Kolloquiums beleuchtete den möglichen Kulturwandel durch KI und betrachtete die Erschließung aus einer erweiterten Perspektive. Hierfür bildeten die Keynote, die Fachbeiträge und der Vortrag der Arbeitsstelle für Standardisierung zur Rolle der klassischen Informationsversorgung in Zeiten von KI die inhaltliche Grundlage für die anschließende Diskussion. Zum Einstieg in die Diskussion griff das Panel die wichtigsten Themen und Fragen aus dem Verlauf der Veranstaltung auf. Das Publikum wurde rasch eingebunden und trug seinerseits mit Fragen und Statements zur Lebendigkeit des Austauschs bei.

Die hier vorliegende Publikation dokumentiert die Beiträge des Kolloquiums in Form von kurzen Texten und verlinkt auf die Präsentationen auf den jeweiligen Servern. Ergänzt wird die Publikation durch eine knappe Zusammenfassung der Paneldiskussion, sowie Poster und Artikel rund um die Themen des Kolloquium selbst, die den zeitlichen Rahmen des Programms jedoch gesprengt hätten.

Photographische Eindrücke aus dem Vortragssaal der Deutschen Nationalbibliothek in Frankfurt



Dorian Grosch (SBB-SPK)

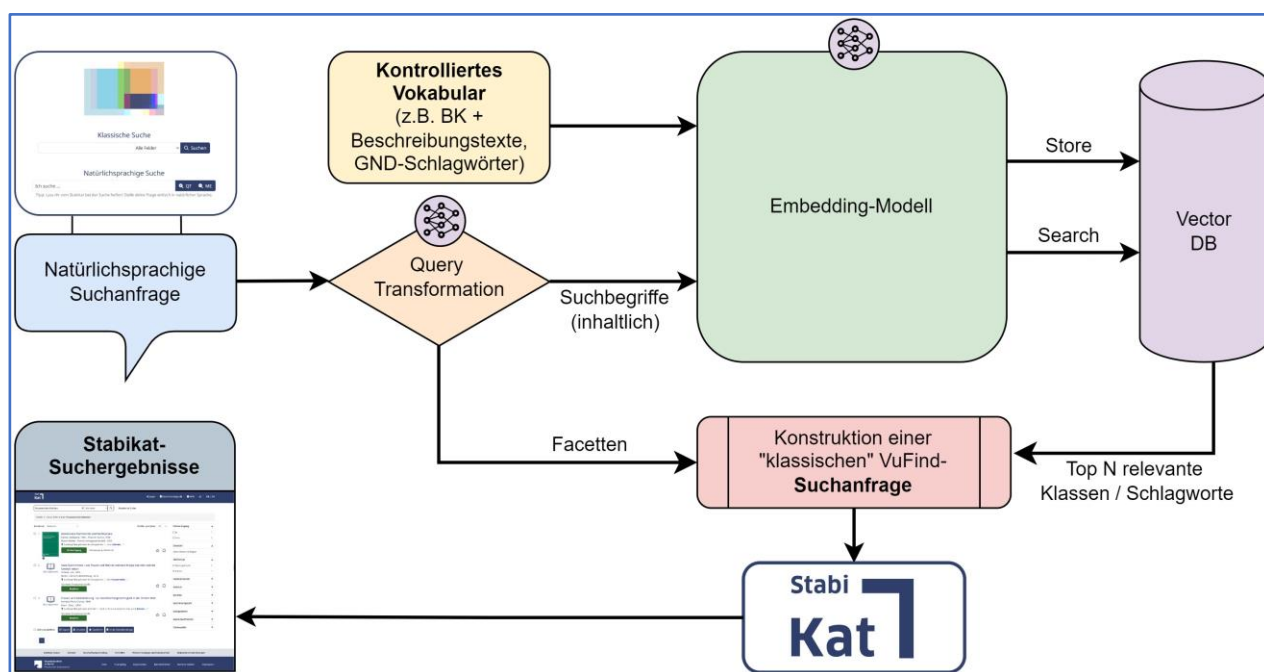
Automatische Rechercheberatung mit LLMs und Text-Embeddings

Vorgestellt wird eine experimentelle Suchmodalität für VuFind-Systeme, welche es den Nutzenden ermöglicht, Rechercheanfragen in natürlicher Sprache zu stellen, statt klassisch nach Stichworten zu suchen.

Der technische Ansatz basiert auf einer sog. Query-Transformation, die eine automatische Rechercheberatung ausführt. Natürlichsprachige Anfragen werden dabei auf Suchfacetten wie Medientyp, Sprache oder Publikationszeitraum sowie auf inhaltliche Suchbegriffe wie BK-Klassen und GND-Schlagwörter abgebildet. Das System nutzt vortrainierte Sprachmodelle (LLMs) und Text-Embeddings (TEs) und vermeidet somit aufwändige ML-Trainingschritte, was eine kosteneffiziente Implementierung ermöglicht.

Die Architektur folgt einem „Plug-In“-Prinzip, bei dem die Rechercheberatung dem Katalog vorgeschaltet wird. Natürlichsprachige Anfragen werden in VuFind-Such-URLs transformiert, wobei eine semantische Vektorsuche in den normierten Begriffen zum Einsatz kommt. Durch kontrollierte Generierung mittels LLMs werden dabei „Halluzinationen“ vermieden und präzise Suchergebnisse im regulären Kataloginterface präsentiert – von wo aus die Recherche fortgeführt werden kann.

Link zum Vortrag: <https://zenodo.org/records/17989697>



Dr. Argie Kasprzik (ZBW)

KI mit Augenmaß

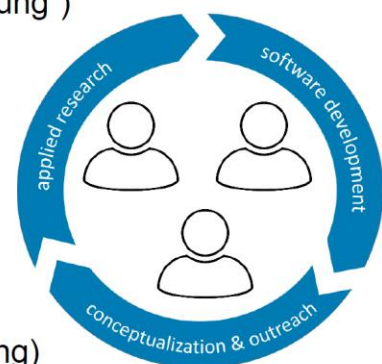
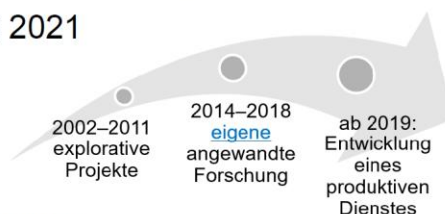
Bericht zum aktuellen Stand der Automatisierung der Inhaltsererschließung (AutoSE) an der ZBW

An der ZBW wurde seit 2014 eigene angewandte Forschung zur Entwicklung von Machine-Learning-Modellen für die Automatisierung der Inhaltsererschließung betrieben. Seit 2021 läuft nun auch ein produktiver Dienst namens AutoSE auf der Basis dieser Modelle, mit der von der Finnischen Nationalbibliothek angestoßenen Open-Source-Plattform Annif als Kernkomponente. Der AutoSE-Dienst wird kontinuierlich weiterentwickelt und auf dem neusten Stand gehalten, sowohl im Hinblick auf den State of the Art im Bereich Softwarearchitektur als auch auf die Methoden. Jüngste Aktivitäten des Teams beziehen sich auf das Integrieren von Transformermodellen in das Modellportfolio von Annif, um diese ebenfalls produktiv nutzbar zu machen. Ein bleibendes Thema ist das begleitende Qualitätssicherungskonzept, sowohl mit automatisierten Herangehensweisen als auch durch das Einbeziehen menschlicher Expertise bei einer gleichzeitigen schrittweisen Rollenverschiebung, zum Beispiel durch Active Learning.

Link zum Vortrag: <https://doi.org/10.5281/zenodo.17941232>

AutoSE: Der Weg zum verstetigten Forschungstransfer

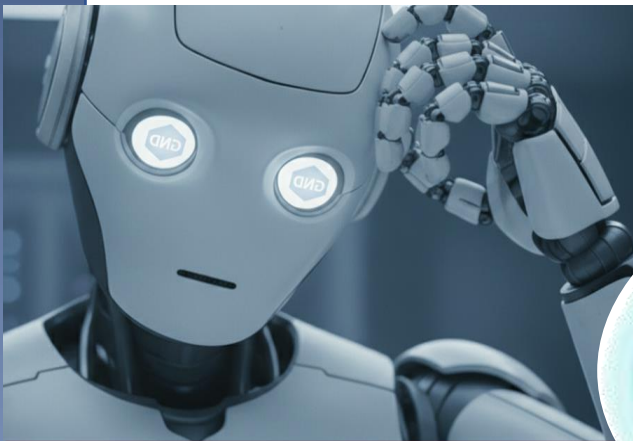
- Vorläufer: Projekt AutoIndex (bis 2018) – **angewandte Forschung** & Prototypen
- ab 2019: **AutoSE** („Automatisierung der Sacherschließung“)
- **produktiver Dienst** seit 2021
- **Open Source!**
- fortlaufend:
Evaluierung und **Integration neuer Entwicklungen aus der KI-Forschung**
(einschließlich unserer eigenen angewandten Forschung)



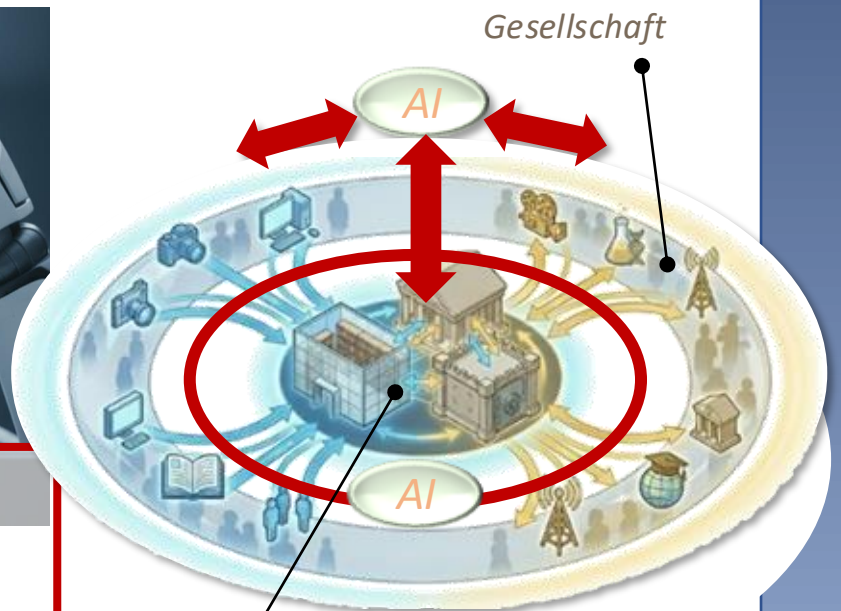
Jürgen Kett (DNB)

Dringend benötigt oder obsolet?

Welche Rolle kann die klassische Informationsversorgung über bibliografische Daten und Normdaten in Zeiten von KI spielen?



- ! KI-generierte Ressourcen, „AI Slop“ und Desinformation
- ! Big-Tech-Dienste als abschirmende Vermittlungsschicht
- ! Veränderte Informationskompetenz
- ! Polarisierung und Vertrauensverlust



Bibliotheken, Archive, Museen als aktives kulturelles Gedächtnis

- ☺ Normdaten als prüfbare Referenzanker für Trusted Data Hubs und KI-Apps.
- ☺ RDA DACH als Qualitätsgarant auch für KI-Agenten.
- ☺ Culturegraph als kooperativer QS-Monitor für die Daten im STA-Netzwerk
- ☺ Der STA als stärkende Struktur für eine selbstbewusste, vertrauenswürdige und kooperative Zukunft mit KI.



Verantwortungsvoll mit dem Menschen im Mittelpunkt

Link zum Vortrag: <https://zenodo.org/records/18314256>

Prof. Dr. Sascha Koch (Jade Hochschule in Oldenburg)

KI in Museen

Museen bewahren nicht nur wertvolle Objekte, sie sind auch Orte des Wissens. Doch ein Großteil dieses Wissens bleibt bislang im Verborgenen: Viele Sammlungen sind nur unvollständig dokumentiert, Herkunftsgeschichten sind schwer nachzuvollziehen und die vorhandenen Daten sind häufig nicht miteinander verknüpft. Die Bereitstellung von Metadaten aus Museen für die Forschung erfährt durch den Einsatz von KI eine transformative Entwicklung. Metadaten, die Informationen über Objekte in Museumssammlungen enthalten, können von KI-Methoden analysiert werden, um tiefgreifende Einblicke und Verbindungen zwischen verschiedenen Artefakten zu generieren. Diese automatisierte Analyse beschleunigt nicht nur den Forschungsprozess erheblich, sondern ermöglicht auch die Entdeckung verborgener Muster und Zusammenhänge, die für menschliche Forscher*innen möglicherweise nur schwer oder nur mit ungleich größerem Aufwand zu erkennen wären. Darüber hinaus trägt die präzise Kategorisierung und Verknüpfung von Metadaten durch KI dazu bei, die Effizienz von Forschungsprojekten zu steigern und den Wissensaustausch innerhalb der wissenschaftlichen Gemeinschaft zu fördern. Die Digitalisierung und intelligente Aufbereitung von Metadaten eröffnen somit neue Horizonte für die Museumsforschung und tragen dazu bei, das Verständnis zu kulturellem Erbe durch innovative Technologien zu vertiefen.

Link zum Vortrag: <http://bit.ly/45F2D23>



Bildnachweis: Jade HS /Fokko Mittelstaedt – KI generiert

Elisabeth Moedden (DNB)

KI in Bibliotheken

Am Ende zählen die Ergebnisse: das DNB-KI-Projekt

Der Vortrag gibt einen kompakten Überblick über die zentralen Ergebnisse des KI-Projekts der Deutschen Nationalbibliothek (2021–2025) zur automatischen Vergabe von GND-Schlagwörtern für deutschsprachige Online-Publikationen. Neben klassischen Extreme-Multi-Label-Classification-Verfahren wird gezeigt, dass transformerbasierte Ansätze wie XRTransformer das Projektziel (F1-Score > 0,4) bereits erreichen und damit die Qualität der automatischen Erschließung messbar steigern. Vor dem Hintergrund der zunehmenden Verfügbarkeit großer Open-Weight-Sprachmodelle wurde zudem die Schlagwortvergabe mit LLMs untersucht. Da LLMs die GND nicht explizit kennen, wurde ein zweistufiger Workflow gewählt: (1) Generierung freier Schlagwortvorschläge per Few-Shot-Prompting und (2) Mapping auf GND-Konzepte mittels Embeddings. Darauf aufbauend wurden drei Verfahren entwickelt: Embedding-basiertes Matching (EBM) als semantisch robustes, dem lexikalischen Ansatz ähnliches Mapping, ein rechenintensiveres LLM-Ensemble zur Kombination mehrerer Modelle und Prompts sowie KIFSPrompt als ressourcensparende, RAG-ähnliche Variante mit intelligenter Beispielauswahl. Die Ergebnisse deuten darauf hin, dass insbesondere bei seltenen oder untrainierten GND-Konzepten zusätzliche Potenziale gegenüber rein klassisch trainierten Modellen bestehen. Abschließend betont der Vortrag die zentrale Rolle der intellektuellen Erschließung für Trainingsbeispiele und Evaluationen und skizziert die Weiterentwicklung modularer Ansätze sowie die Übertragung auf weitere Erschließungsaufgaben.

Link zum Vortrag: <https://zenodo.org/records/18299184>

6

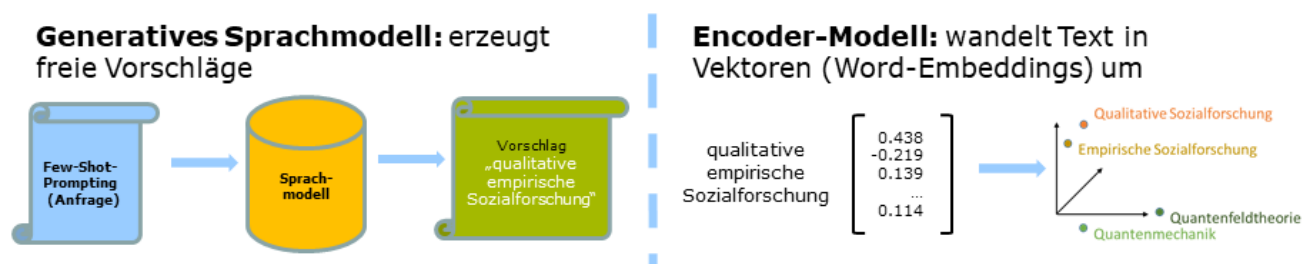
GND-Schlagwortvergabe mit LLMs



Problem: LLMs kennen die GND nicht

Grundprinzip:

1. Generierung freier Schlagwortvorschläge mit einem generativen Sprachmodell
2. Erzeugung eines Mapping auf die GND mithilfe eines (kleineren) Encoder-Modells



Clemens Neudecker (SBB-SPK)

Einblicke in die Anwendung von KI an der Stabi Berlin im Kontext des Projekts „Mensch.Maschine.Kultur“

Der Beitrag durchläuft die wesentlichen Ergebnisse der vier Teilprojekte des BKM-geförderten Forschungsprojekts „Mensch.Maschine.Kultur“ an der Stabi Berlin als Stationen einer datengetriebenen Forschungsreise: Den Beginn machen (1) Intelligente Verfahren für die generische Dokumentenanalyse, die mit hoher Qualität und Präzision Layout-, Text- und Bildinhalte aus digitalisierten Dokumenten extrahieren. Anschließend kommen auf multimodaler KI basierende (2) Verfahren für die Bildanalyse und Bildähnlichkeitssuche zum Einsatz, um in Dokumenten enthaltene Abbildungen zu erschließen und durchsuchbar zu machen. Weiterhin werden Volltexte und Metadaten für die (3) KI-unterstützte Inhaltsanalyse und Sacherschließung genutzt und verschiedene Anif-Modelle trainiert sowie evaluiert. Abschließend erfolgt die (4) Kuratierung und Bereitstellung aller produzierten Werkzeuge, Datensätze und Modelle, wobei sich an Leitfäden und Empfehlungen orientiert wird, die speziell entwickelt wurden um den ethischen, rechtlichen und sozialen Implikationen im Umgang mit KI und Kulturdaten Rechnung zu tragen. Alle Ergebnisse sind frei zugänglich über die Projektwebseite <https://mmk.sbb.berlin/>



Bild: Max Gruber / Better Images of AI / Ceci n'est pas une banane / CC-BY 4.0

Gerald Steilen (GBV)

Das Ende der Erschließungslücke? KI-Technologien für moderne Discovery-Systeme

Wie können Bibliotheken die Erschließungslücke überwinden, wenn nur 23% der bibliographischen Daten über konventionelle Prozesse zugänglich sind? Dieser Vortrag zeigt, wie KI-gestützte semantische Suchverfahren – konkret: mehrsprachige Text-Embeddings und Mixture-of-Experts-Modelle – eine skalierbare Lösung bieten können.

Am Beispiel des Discovery-Systems Lukida an der TU Clausthal (produktiv seit Ostern 2025) wird demonstriert, wie KI-generierte Empfehlungen 100% des Bestands abdecken können, ohne generative Halluzinationen zu riskieren. Die Präsentation adressiert drei zentrale Aspekte: die technische Umsetzung im GBV-Rechenzentrum unter Wahrung der Datensouveränität, die Ressourceneffizienz moderner Embedding-Modelle im Kontext von Green-IT sowie den notwendigen Kulturwandel von „perfekter Erschließung für wenige“ zu „guter Erschließung für alle“.

Der Beitrag plädiert für einen hybriden Ansatz, bei dem KI skaliert und menschliche Expertise kuratiert – und stellt offene Fragen zu Governance, Metadatenstandards und nachhaltiger Finanzierung zur Diskussion.

Link zum Vortrag: <https://zenodo.org/records/17813523>

VZG Verbundzentrale des GBV
STA-Kolloquium "In Zukunft mit KI" 10.12.2025

KI passt sich an Nutzende an
🎓 Verbesserungen + Messbare Erfolge

Drei zentrale Verbesserungen:

- 🧠 Mentale Modelle:
Semantische Suche statt Kataloglogik
- 🔍 Kontextverständnis:
Mehrdeutige Begriffe werden im Kontext verstanden
- 🌐 Sprachübergreifend:
Suchanfragen in einer Sprache finden relevante Treffer in ~100 anderen Sprachen

Messbare Erfolge:

- ✅ Empfehlungen genutzt: 23% aller Recherchen
- ✅ Median-Nutzung: erste 4 Empfehlungen
- ✅ Scrolltiefe: 20% der Klicks auf Positionen 13-50

Empfehlungen: Kumulierte Klickverteilung
Klickverteilung nach Position

Position	Kumulierter Klickanteil (%)
1	20
5	55
10	75
15	85
20	90
25	92
30	94
35	95
40	96
45	97
50	98

Dr. Tomasz Stompór (KOBV)

FAN - Fully Algorithmic Librarian

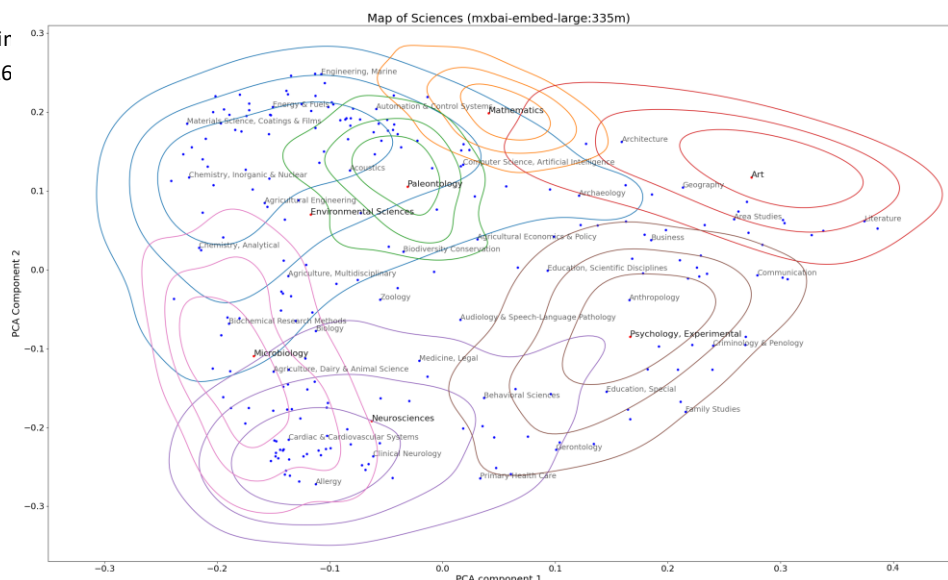
Erfahrungen mit KI auf großen Publikationsdaten

Die Verbundzentrale des Kooperativen Bibliotheksverbundes Berlin-Brandenburg zeichnet sich durch die Besonderheit aus, dass sie in einem landeseigenen außeruniversitären mathematischen Forschungszentrum, dem Zuse-Institut Berlin, integriert ist. In diesem institutionellen Kontext ist das Projekt „Fully Algorithmic Librarian (FAN)“ angesiedelt. Es ist interdisziplinär ausgerichtet und verknüpft Methoden der Mathematik und der Bibliotheks- und Informationswissenschaft. Mithilfe eines großen Zitationsgraphen werden dabei algorithmische Methoden für bibliometrische Fragestellungen erprobt, die auf die Entwicklung von forschungsunterstützenden Services abzielen. Dazu werden groß angelegte Zitationsnetzwerke mit bis zu 119 Millionen Datensätzen und 1,4 Milliarden Zitaten analysiert, die aus den Datenbanken Web of Science und OpenAlex gewonnen wurden. In diesem Prozess werden mathematische Methoden entwickelt, die auf interdisziplinäre Zitationsgraphen angewendet werden können. Ein Beispiel dafür sind algorithmische Clustering-Methoden, mit deren Hilfe die bereits in den Datensätzen vorliegenden Klassifikationen überprüft werden können. Parallel dazu wird diese Methode der automatisierten Klassifizierung durch die Auswertung von Abstracts von Zeitschriftenartikeln unter Verwendung von Embedding-Modellen begleitet. Mögliche Anwendungsszenarien dieser Untersuchungen sind: die automatische Klassifikation von Publikationen, die Visualisierung interdisziplinärer Netzwerke, die Erkennung emergenter Forschungsthemen, die thematische Suche, die Empfehlung von Reviewern und Co-Autor:innen sowie die Offenlegung von blinden Flecken in den Daten (Was ist nicht repräsentiert?).

Link zum Vortrag: <https://zenodo.org/records/18326460>

Das Projekt wird gefördert von
vom 01.11.23 bis 31.10.2026

*Kartierung
eines
Datensatzes
von Web of
Science mit
56 Millionen
Publikationen
anhand
LLM-
Embeddings
der
Abstracts*



STIIV“

Kunt, Tim, et al. "Mapping the Web of Science, a large-scale graph and text-based dataset with LLM embeddings." *ZIB Report* 25-11, July 2025. *opus4.kobv.de*, <https://doi.org/10.12752/10064>

Lizenzhinweis: [CC BY](https://creativecommons.org/licenses/by/4.0/)

In Zukunft mit KI

Was folgt daraus für den Standardisierungsausschuss?

Ein Resümee der Diskussion

Nach einer beeindruckenden Leistungsschau der an den Mitgliedseinrichtungen des Standardisierungsausschusses (STA) bereits laufenden oder im Aufbau befindlichen Workflows und Anwendungen, jeweils auf unterschiedlicher Weise KI-basiert, bildete der [Vortrag von Jürgen Kett](#), Leiter der Arbeitsstelle für Standardisierung, die Grundlage für die anschließende Diskussion.

Der STA entstand 2000 als kooperative Antwort auf den digitalen Wandel und hat mit der Internationalisierung, der Einführung gemeinsamer Standards und dem Aufbau der GND maßgeblich zum Aufbau eines spartenübergreifenden Datenökosystems beigetragen. Mit dem zunehmenden Einsatz von KI verändern sich Arbeitsprozesse grundlegend: Automatisierte Erschließung, wachsende Datenmengen und der Mensch als „Human in the Loop“ prägen die Zukunft. Umso wichtiger sind Standards wie RDA DACH, die eine gemeinsame Vorstellung davon definieren, was einen guten, vertrauenswürdigen Datensatz ausmacht. Insbesondere aber können Normdaten eine noch zentralere Rolle spielen – als verlässliche Referenz, als Grundlage für erklärbare und überprüfbare KI sowie als Mittel gegen Verzerrungen und Halluzinationen. So skizzierte Jürgen Kett gleichermaßen die bestehenden Grundlagen wie auch die Herausforderungen des STA in seinem Vortrag und eröffnete damit die Debatte.

Auf dem Podium stellten sich Reinhard Altenhöner (SBB-SPK), Frank Scholze (DNB), Dr. Silke Schomburg (hbz), Regine Stein (VZG) und moderiert durch Jürgen Kett (DNB) der Herausforderung, zu umreißen, welche Rolle der STA im Zeitalter Künstlicher Intelligenz spielen könne. Das anwesende Publikum nahm regen Anteil an der aufkommenden Debatte. Besonders deutlich wurde, dass der STA sich noch stärker als bisher als einen gestaltenden Zusammenschluss führender Organisationen der Wissensorganisationen im deutschsprachigen Raum verstehen und präsentieren sollte.

Der Kulturwandel durch Künstliche Intelligenz fordert die bestehende und durch die Öffentlichkeit regulierte sowie geförderte Infrastruktur heraus. Daher wurde die Notwendigkeit betont, sich als Gremium dieser Aufgabe gemeinsam zu stellen. Das meine nicht allein eine Erneuerung der Rollendefinition des STA, sondern konkrete Maßnahmen wie die Erarbeitung attraktiver Tätigkeitsprofile, den offenen Austausch über Tools und Anwendungen, die im Kreis der Mitglieder entwickelt werden, die Entwicklung von Qualität sichernden Workflows, die auch Datenmanipulationen standhalten, bis hin zur Harmonisierung und Vereinigung von Geschäftsgängen wie die der Formal- und Inhaltserschließung. Letztlich sollte der STA in einer gemeinsamen Stellungnahme eine Richtschnur zum Umgang mit Künstlicher Intelligenz verabschieden, die wegweisend für die KI-gestützte Digitaltransformation des GLAM-Bereichs und der humanistischen Wissenschaft sein könnte.

Zusammenfassend: Künftige Aufgabe des STA sei es, das gemeinsame Datenökosystem und das Vertrauen der Gesellschaft in sein Netzwerk zu stärken. Denn die Produkte des STA und die Zusammenarbeit zwischen den Partnern können durch den Einsatz der neuen Technologien enorm profitieren, gleichzeitig gelte es, sie für die Nutzung durch KI zu optimieren. Nur durch Zusammenarbeit, Standards und qualitätsgesicherte Daten bliebe verlässliches Wissen auch in der Zukunft zugänglich.

In Zukunft mit KI

Einige Statements aus der Diskussion


- Zum Stichwort „Bedeutungsverlust“: Wir müssen uns Raum geben, für die Formgebung der neuen eigenen Aufgabe.
- Zum Stichwort „human in the loop“: Resterampe oder hybride Intelligenz? Wir brauchen attraktive Tätigkeitsprofile in der Bibliothek.
- Zum Stichwort „Qualität“. Gemeinsam an der Homogenisierung der Qualität in den Datenlieferungen arbeiten. Culturegraph als ein Ansatz für den Mehrwert der Interoperabilität von Ressourcen vorstellen. Das Monitoring der Daten schützt auch vor Datenmanipulationen.
- Zum Stichwort „Standards“: Wir sollten uns endgültig von der mentalen Karteikarte trennen, dabei jedoch zugleich Qualitätsstandards für Metadaten bewerben. Wie sieht die Perspektive für MARC21 aus?
- Zum Stichwort „Zusammenarbeit“: Der STA sollte eine gemeinsame Policy zu KI generierten Inhalten entwickeln.
- Zum Stichwort „Zusammenarbeit“: Die Geschäftsgänge der Formal-, Inhalts- und Normdatenerschließung vereinter behandeln.
- Zum Stichwort „Zusammenarbeit“: Mehr Standardisierung und Zugänglichkeit der im STA Netzwerk generierten Werkzeuge.
- Zum Stichwort „Zusammenarbeit“: Das bedeutet intensivere Zusammenarbeit über die Gremienarbeit hinaus.

Es folgen Beiträge der Partner, die zum Standardisierungsausschuss, zu seinen Themen und Produkten sowie zum Thema des Kolloquiums Bezug haben.

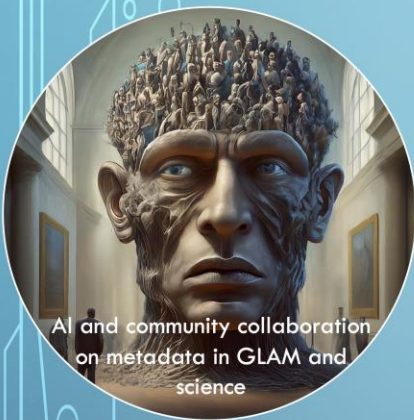
Barbara Fischer (PID Network Deutschland)

Empowering AI Knowledge Management, 2024

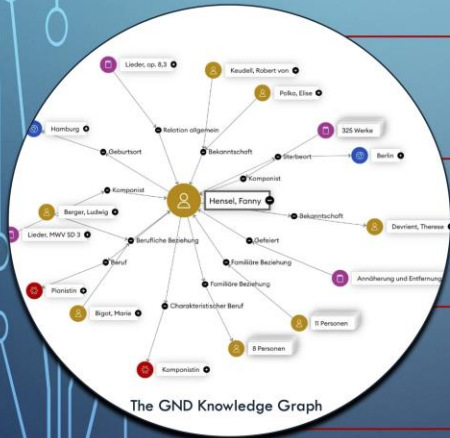
EMPOWERING AI KNOWLEDGE MANAGEMENT



A COMMUNITY-ORGANIZING APPROACH TO ENHANCE FIDELITY AND QUALITY THROUGH AUTHORITY FILE USE IN METADATA



AI helps creating metadata automatically. Authority data minimizes the risks of AI hallucination and increases precision and recall results in metadata based operations. Even more, authority data may embrace the the diversity of GLAM and science communities. Thus, the German National Library and its partners further an participative infrastructure that empowers a broad range of diverse communities to use, contribute and edit the Integrated Authority File (GND) in German-speaking countries.



informing and building community

tools for exploring

community empowerment

Interoperable APIs

Organizing Communities

DFG Deutsche Forschungsgemeinschaft

DataCite

DEUTSCHE NATIONAL BIBLIOTHEK

HELMHOLTZ Open Science

TIB UNIVERSITÄT BIELEFELD Universitätsbibliothek

PID NETWORK DEUTSCHLAND

Stabi Berlin



Natürlichsprachige Suche im Stabikat

Automatische Rechercheberatung mit LLMs und Text-Embeddings

Medientyp

Mediensprache

Ich suche Bücher und Artikel in Englisch und Spanisch aus den letzten 40 Jahren über Frauenrechte in Bolivien.

Publikationszeitraum

Inhaltliche Aspekte der Suche

Das Projekt "Natürlichsprachige Suche im Stabikat" entwickelt eine experimentelle Suchmodalität, welche es den Nutzenden ermöglicht, Rechercheanfragen in natürlicher Sprache zu stellen, statt klassisch nach Stichworten zu suchen.

Oben: Ein Beispiel für eine natürlichsprachige Suchanfrage. Es sind verschiedene formale und inhaltliche Aspekte der Recherche zu erkennen.

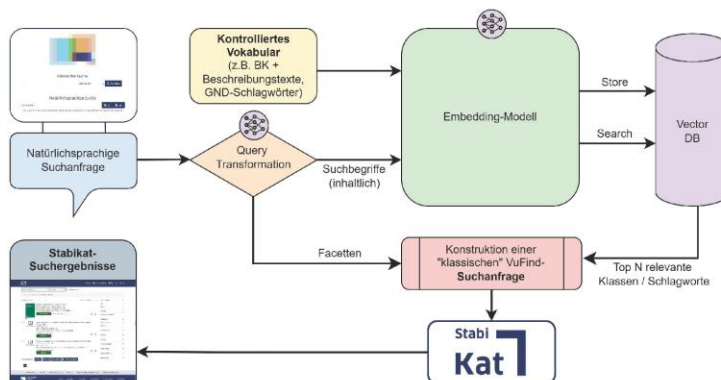
Unten: Vereinfachte Architektur des Systems zur natürlichsprachigen Suche. Die Anfrage wird durch mehrere Schritte in eine Stabikat-Such-URL (nach VuFind-Syntax) umgeformt.

Suchtechnologie im Wandel

ChatGPT & co. verändern die Ansprüche von Nutzenden an die Literaturrecherche. Mit der **natürlichsprachigen Suche** wird dafür im Stabikat eine neuartige Suchmodalität implementiert: Nutzende stellen Anfragen einfach in natürlicher Sprache, und mittels KI-Technologien werden passende Filter, Klassen oder Schlagworte für die Recherche vorgeschlagen.

Der technische Ansatz basiert auf einer sog. **Query-Transformation**, die eine automatische Rechercheberatung ausführt. Natürlichsprachige Anfragen werden dabei auf **Suchfacetten** wie Medientyp, Sprache oder Publikationszeitraum sowie auf **inhaltliche Suchbegriffe** wie BK-Klassen und GND-Schlagwörter abgebildet. Das System nutzt vortrainierte Sprachmodelle (LLMs) und Text-Embeddings (TEs) und vermeidet somit aufwändige ML-Trainingschritte, was eine kosteneffiziente Implementierung ermöglicht.

Die Architektur folgt einem **"Plug-In"-Prinzip**, bei dem die Rechercheberatung dem Stabikat vorgeschaltet wird. Natürlichsprachige Anfragen werden in **Stabikat-Such-URLs** (nach VuFind-Syntax) transformiert, wobei eine semantische Vektorsuche in den normierten Begriffen zum Einsatz kommt. Durch kontrollierte Generierung mittels LLMs werden dabei "Halluzinationen" vermieden und präzise **Suchergebnisse im regulären Stabikat-Interface** präsentiert - von wo aus die Recherche fortgeführt werden kann.



Oben: Die kontrollierten Vokabulare werden speziell aufbereitet, damit eine **semantische Vektorsuche** möglich wird. Diese wird bei der Verarbeitung einer Anfrage ausgeführt, um relevante Klassen oder Schlagworte auszuwählen.

Unten: Beispiel für das **Interface** einer natürlichsprachigen Suche. Die Anfrage wird in einen Suchschlitz eingegeben, in der Suche erscheinen danach die **vorgeschlagenen Suchfilter**.

Für die **semantische Suche** wurden die Basis-klassifikation und GND-Sachbegriffe strukturiert aufbereitet und mit natürlichsprachigen Beschreibungen angereichert. Diese **Datenaufbereitung** bildet die Grundlage für die präzise Zuordnung von Suchanfragen zu normierten Begriffen.

Natürlichsprachige Suche

Ich suche Bücher und Artikel ... über Frauenrechte in Bolivien

Tipp: Lass dir vom Stabikat bei der Suche helfen! Stelle deine Frage einfach in natürlicher Sprache.

Basierend auf deiner Anfrage hat der Stabikat folgende Suchfilter generiert:

Sprache: Englisch ODER Spanisch Medientyp: Aufsatz ODER Buch

Schlagwörter: Frauenbewegung ODER Frauenpolitik ODER Bolivien Erscheinungsjahr: 1985 - 2025

Vorteile des Ansatzes:

- Keine Veränderung am Katalog-Backend nötig
- Übertragbar auf andere VuFind-Systeme
- Erweiterbar durch weitere Suchlogik, kontrollierte Vokabulare und Klassifikationen

Ansprechpartner zum Projekt:

Dorian Grosch
dorian.grosch@sbb.spk-berlin.de
Wissenschaftlicher Mitarbeiter
Staatsbibliothek zu Berlin



Andrea Hemmer (DNB)

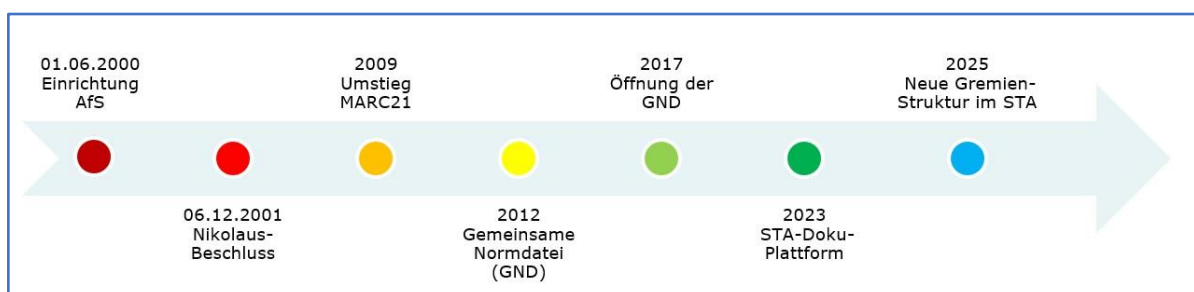
Zwischen Bewahren und Bewegen – 25 Jahre Arbeitsstelle für Standardisierung

ein Blogpost

In diesem Jahr jährt sich die Gründung der Arbeitsstelle für Standardisierung (AfS) zum 25. Mal. Dies ist ein guter Anlass Revue passieren zu lassen, zu betrachten, welche Schwerpunkte die Arbeit prägten und zu sehen, was in der Zeitspanne erreicht wurde.

Der DNB-Blogpost zum Jubiläum der Geschäftsstelle thematisiert den Anlass der zur Gründung führte, den legendären „Nikolaus-Beschluss“ zur Internationalisierung der Arbeit, das auch im Kolloquium thematisierte Format MARC-21, den Umstieg auf das internationale Regelwerk RDA, die Öffnung der Gemeinsamen Normdatei im Jahr 2017 und resümiert die heute bestehenden Infrastrukturangebote zur Integration neuer Communities, die mit ihrer Expertise und ihren Daten zur Umsetzung der Linked-Open-Data-Vision beitragen.

Den vollständigen Text finden Sie hier: <https://blog.dnb.de/zwischen-bewahren-und-bewegen/>



Christoph Poley, Maximilian Kähler et al. (DNB) 2025

Automatic Subject Cataloguing at the German National Library

The German National Library (DNB) began developing solutions for automatic subject cataloguing 15 years ago. The main reason for this was the huge and ever-growing number of digital media works that needed to be indexed. Today, the DNB uses open source algorithms and frameworks to assign various types of thematic meta information in this way.

This practice paper provides a deeper insight into automatic subject cataloguing at the DNB. We look at the data and vocabularies used as well as at the different methods and approaches. The vocabulary for classification is based on the Dewey Decimal Classification (DDC). For verbal subject indexing we use the German Integrated Authority File (GND).

The use case of automatic classification is divided into the assignment of DDC Subject Categories and DDC Short Numbers. Due to the large size of the GND vocabulary, the use case of automatic indexing is an extreme multi-label classification (XMLC) problem. A brief report is given about the construction and the performance of our models.

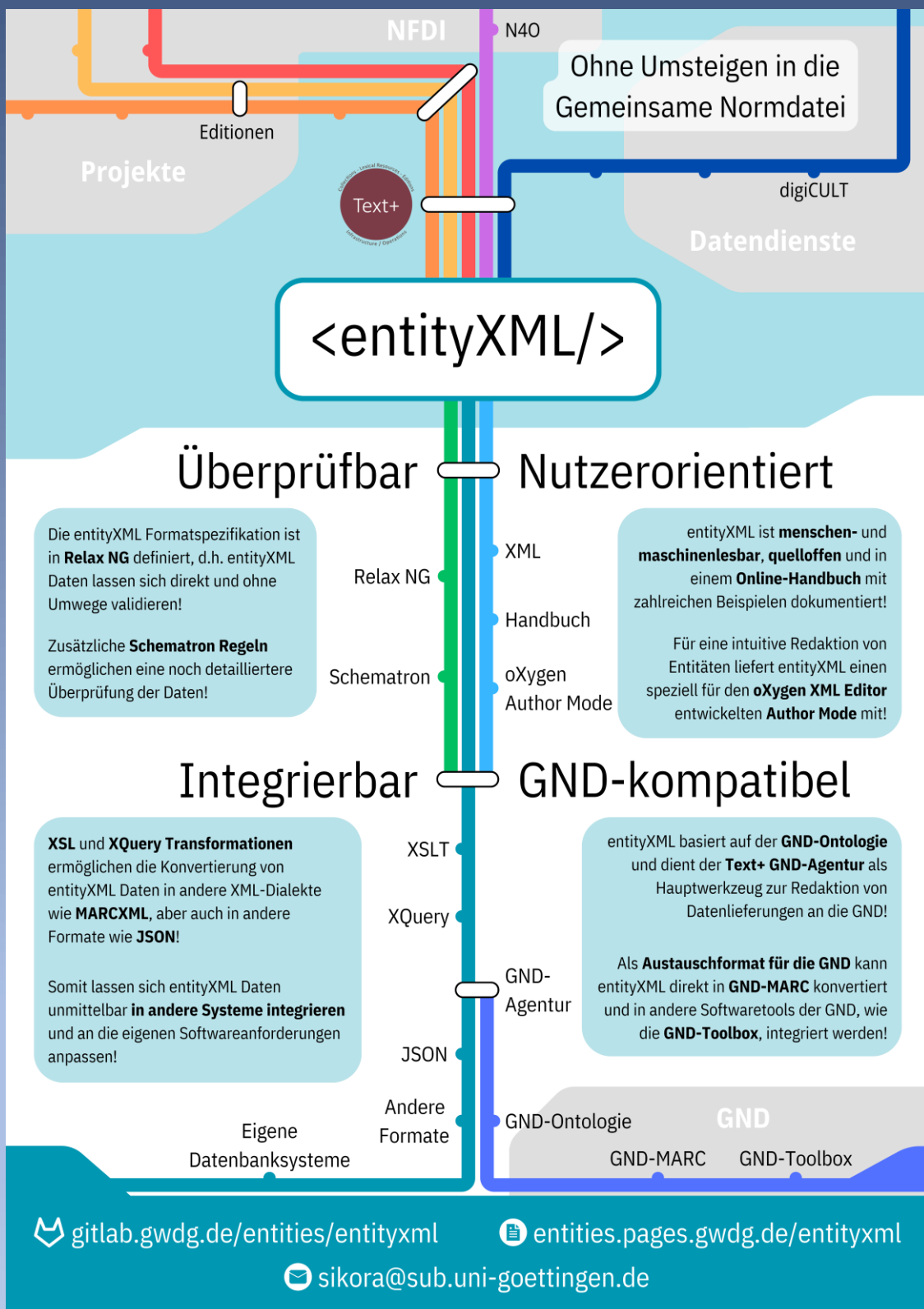
Based on these use cases, we present some implementation aspects of our “subject cataloguing machine” EMa, the environment for automatic subject cataloguing in productive use. We point out the basic feature set and provide a high-level introduction of the productive EMa system. The modular design of the EMa software architecture with the open source software Annif as a central toolkit is described.

The development of EMa is an ongoing task at the DNB. It requires continuous development and maintenance, technological and human resources. Applied research activities in the DNB's AI project are closely related to the EMa ensuring that relevant scientific findings get integrated into its development.

Link to full article: <https://liberquarterly.eu/article/view/19422>

Uwe Sikora (NFDI Text+)

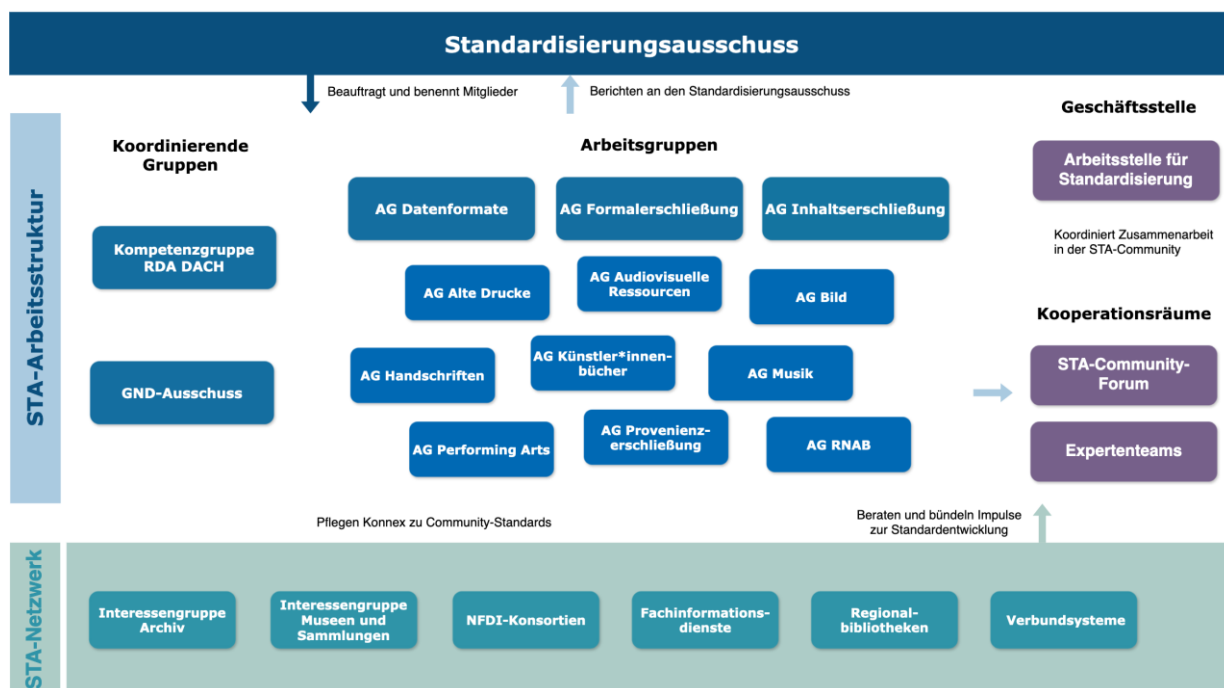
Der Entity-xml Workflow, 2024



Der Standardisierungsausschuss und seine Gremien

Der Standardisierungsausschuss (STA) ist ein kooperativer Zusammenschluss von Organisationen der Kultur und Wissenschaft im deutschsprachigen Raum, mit der Zielsetzung, den Einsatz einheitlicher Standards für die Erschließung, Schnittstellen und Formate in Bibliotheken sicher zu stellen und die spartenübergreifende Harmonisierung der Erschließung und Datenvernetzung zu fördern. Das Einbringen von Fachanforderungen unter anderem an die GND übernehmen die Arbeitsgruppen des [STA](#). Hierbei handelt es sich um Fachgremien mit Expertise zu Standardisierungsfragen in unterschiedlichen Bereichen, Materialien und Themen. Sie werden durch den STA bei Bedarf eingerichtet, rekrutieren sich aus dem Partnernetzwerk des STA und erarbeiten Entscheidungsvorlagen für Änderungen, Anpassungen, Zuweisungen sowie Präzisierungen der Regelwerke. Neben den klassischen Themenfeldern rund um die Erschließung und den Datenaustausch in Bibliotheken (verantwortet durch die Arbeitsgruppen Formalerschließung, Inhaltserschließung und die Datenformate) existieren auch Spartenübergreifende Arbeitsgruppen zur Erschließung von speziellen Materialien wie z. B. audiovisuelle Medien, Musik, Bilder, Handschriften und Performing Arts. Ergänzend zu den Arbeitsgruppen bieten Interessengruppen die Möglichkeit, die Anliegen ihrer Community im Rahmen der GND zu vertreten. In den Interessengruppen tauschen sich die teilnehmenden Organisationen aus einem Bereich der Kultur und der Wissenschaft zu ihren Bedarfen an die GND und deren Standards aus. Sie sind selbstorganisierte Gruppen.

Link zur weiterführenden Information: https://gnd.network/Webs/gnd/DE/Gestalten/gestalten_node.html



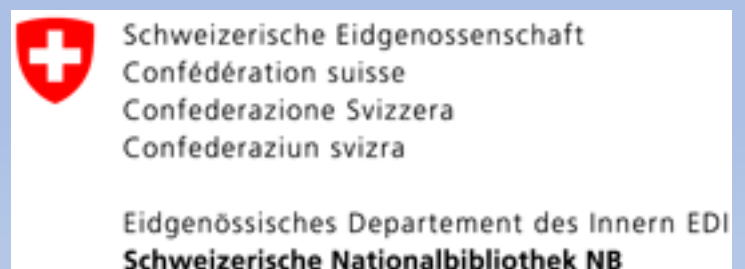
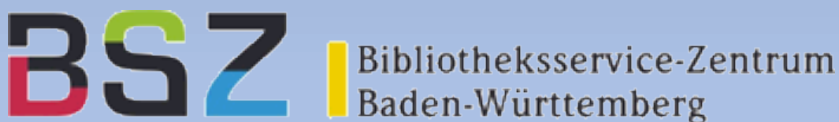
Der Standardisierungsausschuss und seine Gremien, Credit: Afs (DNB) 2024, CC BY SA

Weiterführende Informationen

- Der [Link](#) zum Standardisierungsausschuss
- Der [Link](#) zur STA-Dokumentationsplattform
- Der [Link](#) zur STA-Community
- Der [Link](#) zur GND-Website

Herausgeber:
Arbeitsstelle für Standardisierung an
der Deutschen Nationalbibliothek,
Frankfurt und Leipzig
2026

Wir danken den Förderern



Lizenzhinweis: [CC BY SA](#) für die Gesamtpublikation,
abweichende Lizenzierung der Einzelbestandteile ist möglich.