

Normdatenverknüpfungen aus Culturegraph für die Bildungsforschung (und darüber hinaus)

Marcel Jungmann, DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt am Main
Angela Vorndran, Deutsche Nationalbibliothek, Frankfurt am Main

Christoph Schindler, DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt am Main

Zusammenfassung

Der Fachinformationsdienst (FID) Erziehungswissenschaft und Bildungsforschung hat die Zielsetzung, seine Literaturnachweise im Fachportal Pädagogik systematisch mit Normdaten zu verknüpfen. Ein höheres Maß an Normierung bietet im Bereich der Autorschaft von wissenschaftlichen Publikationen den Vorteil, Personen eindeutig zu identifizieren, größere Vernetzung und Verlässlichkeit der Daten zu erreichen und weitere Informationen über Personen leichter zugänglich zu machen. Für den Aufbau von Personenprofileiten im Fachportal sind Normdatenverknüpfungen ebenfalls wichtig, um automatisiert biografische Angaben übernehmen zu können. Die Deutsche Nationalbibliothek (DNB) setzt seit mehreren Jahren Verfahren zur Disambiguierung von Personen in Titeldaten ein, die auf einer Werkbündelung der Publikationen des umfangreichen Culturegraph-Bestandes (DNB, deutsche und österreichische Bibliotheksverbünde) basieren. Seit Kurzem besteht auch die Möglichkeit, externe Datenbestände punktuell in die Werkbündelung zu integrieren und Personennormdatenverknüpfungen für diese zu ermitteln. Vor dem Hintergrund dieser Zielsetzungen sind die beiden Institutionen dieses Kooperationsprojekt eingegangen. Für die zentralen Teildatenbestände des Fachportals Pädagogik, der FIS Bildung Literaturdatenbank und einer Version der Datenbank des Education Resources Information Center (ERIC) wurden erfolgreich ca. 365.000 Verknüpfungen zu Personennormdatensätzen der Gemeinsamen Normdatei (GND) in ca. 270.000 Datensätzen identifiziert. Auch Open Researcher and Contributor IDs (ORCID iDs) konnten mithilfe einer Konkordanz zur GND ermittelt werden.

Summary

The Specialised Information Service for Educational Science and Educational Research (FID Erziehungswissenschaft und Bildungsforschung) aims to systematically link its bibliographic records in the Education Research Portal (Fachportal Pädagogik) with authority data. A higher degree of standardization in the field of authorship of scientific publications offers advantages in the unambiguous identification of authors, interconnectedness and reliability of records and entities, and in making further information about individuals more easily accessible. Authority data links are also important for creating author profile pages in the Education Research Portal so that biographical information can be retrieved. The German National Library (DNB) has developed and established processes for disambiguating persons and enriching bibliographic records with links to name authority records for several years. These processes are based on forming work clusters amongst the extensive holdings of Culturegraph (DNB, regional library networks in Germany and Austria). Recently, an opportunity was created to incorporate external datasets on a project basis into the work-clustering process, enabling the identification of the corresponding personal authority records in the German Integrated Authority File (GND). Against the background of these aims both institutions have entered this cooperation. For

the main bibliographical collections of the Education Research Portal, the German Education Index (FIS Bildung Literaturdatenbank), as well as a version of the Education Resources Information Center (ERIC) database, approximately 365,000 links to personal authority records in the GND were successfully identified in 270,000 bibliographical records. Open Researcher and Contributor IDs (ORCID iDs) were also determined by using a mapping of GND control numbers and ORCID iDs.

Schlagwörter: Fachportal Pädagogik; FID; Culturegraph; GND; Normdaten; Persistent identifier; Metadaten; Kataloganreicherung

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/6187>

Autorenidentifikation: Marcel Jungmann, ORCID: [0009-0003-1048-1629](https://orcid.org/0009-0003-1048-1629),
Angela Vorndran, ORCID: [0000-0001-7162-9875](https://orcid.org/0000-0001-7162-9875),
Christoph Schindler, ORCID: [0000-0003-3540-370X](https://orcid.org/0000-0003-3540-370X)

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

1. Einleitung

Die Sichtbarkeit von Autor*innen sowie ihrer Publikationen ist zentral für die Rezeption wissenschaftlicher Arbeiten, die Nachverfolgung akademischer Laufbahnen und die gezielte Recherche nach Forschungsergebnissen. Entsprechend macht es sich der Fachinformationsdienst (FID) Erziehungswissenschaft und Bildungsforschung zur Aufgabe, im Fachportal Pädagogik durch Personenprofilseiten die Sichtbarkeit und Kontextualisierung wissenschaftlicher Publikationen zu optimieren. Diese Profile sollen auf Normdaten basieren und neben biografischen Informationen insbesondere auch eine bestandsübergreifende Darstellung der wissenschaftlichen Outputs von Autor*innen in Form personengebundener Literaturlisten ermöglichen.¹ Abb. 1 zeigt einen Entwurf der zurzeit im Aufbau befindlichen Profilseiten.

1 Zur Rolle von Persistent Identifier (PID) in der Forschungsinfrastruktur: Meadows, Alice; Haak, Laurel L.; Brown, Josh: Persistent Identifiers. The Building Blocks of the Research Information Infrastructure, in: Insights 32, 2019, <https://doi.org/10.1629/uksg.457>.

Fachportal Pädagogik - Marie Musterfrau
<https://www.fachportal-paedagogik.de/personen/marie-musterfrau-250134>

[Startseite](#) > [Personen](#) > Marie Musterfrau

Marie Musterfrau

M. Musterfrau; Musterfrau, Marie; C.M. Musterfrau

Persistent Identifiers:
[GND: 10000001-4](#) [VIAF: 1244235](#) [WikiData: Q1234](#)

Länderbezug
Deutschland Österreich

Schlagwörter/Keywords:
[Hochschulforschung](#) [Hochschullehre](#)

Themengebiete
Hochschulforschung & Hochschuldidaktik; Bildungssoziologie

Beziehung mit Körperschaften

- Organization
- Organization
- Organization

Beruf

Akademischer Grad

Datenquelle: GND
update: xx. xx. xxxx

Biografische Angaben

dt. Psychologin; Dias, Potsdam 1999; Habil.-Schr. 2005; seit Dez. 2005 Lehrstuhl für Empir. Bildungsforschung an der Otto-Friedrich-Univ. Bamberg; Forschungsinteressen: Theorien des Lesens und der Lesekompetenz, Entwicklungsdeterminanten, Metakognition (Entwicklung und Fördermöglichkeiten), Selbstreguliertes Lernen, Verknüpfung motivationaler und kognitiver Prozesse beim Lernen, Methoden von Large Scale Assessments, Kompetenzdiagnostik

Publikationsschlagwörter

Hochschuleystem (8) Hochschultyp (7) Hochschulumwelt (4) Hochschulunterricht (2)
Hochschulverband (2)

Co-Authors

Co-Author1 (4) Co-Author2 (2) Co-Author3 (2) Co-Author4 (1)

Publikationen

1
Engaging faculty in group-level change for institutional transformation. Disrupting inequality and building inclusive academic departments.
Jackson, J. Kael
Monographie | New York, NY: Routledge | 2023

Projekte

Mitarbeit in:

Projekttitel 1 [Link aufs Projekt bei uns](#)
Projekttitel 2 [Link aufs Projekt bei uns](#)
Projekttitel 3 [Link aufs Projekt bei uns](#)

Eingeworben:

Projekttitel 4 [Link aufs Projekt bei uns](#)

Transferaktivitäten

[Link auf Blogbeitrag](#)
[Link auf Zeitschriftenartikel](#)

Veranstaltungen

Person steht in Beziehung mit Veranstaltung:
[Link auf Veranstaltung beim DBS](#)
[Link auf Veranstaltung beim DBS](#)

Qualifikationsarbeiten

[Link auf proHabil](#)

Abb. 1: Entwurf eines Profils mit GND-Daten und Informationen aus Titeldaten, Quelle: Eigene Darstellung

Voraussetzung für solche Profilsseiten ist die Verknüpfung von Literatursatzen mit Normdatensätzen zu Personen. Nur wenn solche Verknüpfungen bestehen, können Literatursätze zuverlässig unter einem Personenprofil gebündelt angezeigt werden. Personen-Identifizier bzw. Normdatensätze sind ebenfalls nötig, um Profile automatisiert mit biografischen Informationen aus der Gemeinsamen Normdatei (GND), oder auch anderen Quellen wie Open Researcher and Contributor ID (ORCID) oder WikiData, zu bestücken. Gerade in einem multidisziplinären, mehrsprachigen und historisch gewachsenen Datenkontext wie dem des Fachportals Pädagogik stellt das Fehlen von Personen-Iden-

tifiern jedoch eine Herausforderung dar.² Hinzu kommt, dass die Datenhaltung des FID nicht direkt an ein bibliothekarisches Verbundsystem angeschlossen ist.

Der vorliegende Artikel greift zentrale Aspekte dieses Themas auf.³ Er beschreibt schwerpunktmäßig den Ansatz und die Durchführung eines Verfahrens zur automatisierten Ermittlung von Personen-Identifiern auf Basis der GND, um Personeneinträge in Titeldaten des FID Erziehungswissenschaft und Bildungsforschung mit diesen anzureichern. Das Projekt fand als Kooperation des DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation und der Deutschen Nationalbibliothek (DNB) statt. Ergänzend wird ein zweiter Ansatz vorgestellt und geprüft: Dieser beruht auf einem Abgleich mittels Digital Object Identifier (DOI) über eine Schnittstelle der DOI-Vergabestelle Crossref. Beide Verfahren zielen darauf ab, möglichst valide und bestandsabdeckende Verknüpfungen zwischen Literatur- und Normdatensätzen herzustellen. Diese Normdaten und Verknüpfungen bilden die technische Grundlage für die Umsetzung der geplanten Personenprofile im Fachportal.

2. Konzeptionelle Vorarbeiten und Ansätze im Fachportal Pädagogik

Mit der Entwicklung von personenbasierten Profilsseiten im Fachportal Pädagogik wird das Ziel verfolgt, an Publikationen beteiligte Personen über individuell adressierbare Informationsseiten zu repräsentieren. Diese individuelle Adressierbarkeit via URL dient dabei der Verlinkung mit weiteren Ressourcen.

Mehrere Fachinformationsdienste bzw. Rechercheportale bieten Personenprofilsseiten an. Dabei gibt es zwei Formen der Ausgestaltung: Einerseits dedizierte Profilsseiten, die über eine feste URL adressierbar sind, andererseits die Anzeige von Personeninformationen in einblendbaren Info-Bereichen („Pop-Ups“). In drei von sechs untersuchten Fällen ist ein GND-Identifier Teil der URL-Struktur. Alle Portale zeigen Informationen aus GND-Sätzen an, die vermutlich über GND-Webschnittstellen eingebunden werden.⁴ Die Verlinkung von Profilsseiten erfolgt in der Regel aus Literaturnachweisen heraus mittels interaktiver Elemente neben dem Namen.

Die Grundlage sowohl für die Anzeige von Personeninformationen als auch für die Generierung personengebundener Literaturlisten ist eine systematische Verknüpfung von Literatur- und Personendaten. In der Datenpraxis des FID Erziehungswissenschaft und Bildungsforschung gestaltet sich dies jedoch schwierig. Die FIS Bildung Literaturdatenbank wird kooperativ von über 30 Einrichtungen erstellt, wobei die zugelieferten Metadatenformate und die Erfassungsstandards der beitragenden Einrichtungen stark variieren.⁵ Zudem erfolgt die Datenhaltung in einer eigenen Datenbankanwendung ohne

2 Oerder, Thomas; Schindler, Christoph; Rittberger, Marc: Vom Fachinformationssystem Bildung zum Fachportal Pädagogik. Die Gestaltung der Digitalisierung der Wissenschaft am Beispiel der Bildungsforschung, in: Bibliotheksdienst 52 (3/4), 2018, S. 196–207, <https://doi.org/10.1515/bd-2018-0027>.

3 Dieser Beitrag bezieht sich auf den Vortrag der Autor*innen mit gleichem Titel am 26.06.2025 beim 9. Bibliothekskongress 2025 (zugleich 113. BiblioCon) in Bremen. Die Vortragsfolien sind abrufbar unter: <https://opus4.kobv.de/opus4-bib-info/front-door/index/index/docId/19789>, Stand: 18.09.2025.

4 Schnittstellen wie „lobid“ und „DNB EntityFacts“ ermöglichen den Abruf von strukturierten Daten für gegebene GND-Identifier.

5 Vgl. Oerder; Schindler; Rittberger: Vom Fachinformationssystem Bildung zum Fachportal Pädagogik, 2018.

Anschluss an einen Verbundkatalog, wo Normdatenverknüpfungen in Titeldaten auf breiter Basis verfügbar wären. Personen-GND-Identifizierer sind daher in der FIS Bildung Literaturdatenbank bislang nur für einen Teil der an Publikationen beteiligten Personeneinträge vorhanden. Zudem sind internationale Quellen wie die Datenbank des „Educational Resources Information Center“ (ERIC) oder Auszüge der Library of Congress im Fachportal integriert, die keine GND-Identifizierer enthalten. Im Fall von ERIC ist festzustellen, dass die im Wissenschaftssystem gängige ORCID zunehmend genutzt wird. Bereits über 80.000 ORCID-Identifizierer mit über 130.000 Dokumentverknüpfungen sind in der Sammlung enthalten.

Aus diesen Voraussetzungen wurden mehrere, sich gegenseitig ergänzende, Anreicherungsstrategien verfolgt:

1. *Konsolidierung bestehender Dokument-Personen-Verknüpfungen* der FIS Bildung Literaturdatenbank (GND), des Open-Access-Repositoriums peDOCS (GND, ORCID) und der ERIC-Datenbank (ORCID) in einer zentralen Datenhaltung.⁶
2. *DOI-basierter Abgleich mit externen Quellen* (z. B. Crossref), um ORCID-Identifizierer zu ermitteln und Publikationen zu verknüpfen (siehe Abschnitt 3).
3. *Abgleich mit großen Datenbeständen wie Verbundkatalogen oder dem DNB-Katalog* zur Anreicherung von Publikationen mit GND-Identifizierern (siehe Abschnitt 4).
4. *Nutzung von GND-ORCID-Konkordanzen* als weitere Möglichkeit aus der DNB-Kooperation (siehe Abschnitt 5).

Zunächst wurden im FID die ersten beiden Ansätze verfolgt. Punkte 3 und 4 wurden in Kooperation mit der DNB bearbeitet und werden in den folgenden Kapiteln ausführlicher beschrieben.

3. DOI-basierter Abgleich mit externen Quellen

Ein erster Ansatz zur Ermittlung von Personen-Dokument-Verknüpfungen bestand in einem automatisierten Abgleich von Literaturdatensätzen mit externen Quellen über DOI. Ziel war es, vorhandene Literaturdatensätze um ORCID-Identifizierer zu ergänzen. Für diesen Abgleich wurden mehrere bibliografische Datenquellen aus dem Fachportal genutzt: die FIS Bildung Literaturdatenbank, die ERIC-Datenbank sowie fachliche Ausschnitte aus dem Datenbestand der einschlägigen früheren Sondersammelgebiete (Online Contents, OLC). Der Abgleich erfolgte über die Programmierschnittstelle (API) von Crossref.⁷ Als Abgleichmerkmal diente der in den Metadaten vorhandene DOI.

Die Ergebnisse in Tab. 1 zeigen sowohl das Potenzial als auch die Grenzen dieses Ansatzes.

6 Die FIS Bildung Software unterstützt auf Schema- und Anwendungsebene nur einen Identifizierer (eine GND-Nummer) pro Person. Die Konsolidierung erfolgt in einer speziellen Datenbank, die sowohl Personen in Titeldaten mit Personen-Identifizierern verknüpft als auch Mappings zwischen Identifizierern realisiert.

7 Crossref ist eine DOI-Vergabestelle, über deren Web-API Metadaten von Publikationen und damit zusammenhängenden Entitäten abgefragt werden können.

Tab. 1: Abgleich von Metadaten-Sammlungen im Fachportal mit Crossref

DQ	# DOI	# API OK	# API NOK	# PUB	# PV	# ORCID	# GND
FIS	65.108	57.773	7.335	3.753	5.646	4.700	1.298
ERIC	333.374	332.413	938	65.421	99.756	66.691	5.790
OLC	44.314	44.258	56	5.428	7.281	5.843	819

Spalte „DQ“ verweist auf die Bezeichnung der Datenquelle im Fachportal Pädagogik, „# DOI“ weist die Anzahl der jeweils enthaltenen DOI aus. Nicht alle DOI waren der Crossref-API bekannt („# API NOK“). Im Erfolgsfall lieferte die API häufiger einen oder mehrere mit der Publikation verknüpfte ORCID-Identifizier für eine oder mehrere beteiligte Personen zurück. „# PUB“ gibt die Anzahl der Publikationen mit mindestens einem ORCID-Identifizier an. Die Gesamtanzahl der Publikation-ORCID-Verknüpfungen gibt Spalte „# PV“ an. „# ORCID“ gibt die Anzahl der distinkten ORCID-Identifizier aus Spalte „# PV“ in der Ergebnismenge an. Die Spalte „# GND“ weist aus, wie viele GND-Identifizier über eine später verfügbar gewordene GND-ORCID-Konkordanz aus den in „# ORCID“ repräsentierten Identifiern abgeleitet werden konnten (siehe Abschnitt 5).

Zum Zeitpunkt des Crossref-Abgleichs war FIS Bildung auf Stand Oktober 2022 mit ca. 990.000 nachgewiesenen Publikationen, in ERIC waren ca. 1,9 Millionen und im OLC-Ausschnitt ca. 170.000 Publikationen nachgewiesen. Auffällig ist die vergleichsweise hohe Anzahl von DOI in der FIS Bildung Literaturdatenbank, die nicht über die Crossref-API aufgelöst werden (ca. 11,3 %). Während sich aus der ERIC-Datenbank im Vergleich zu FIS und OLC deutlich mehr ORCID-Verknüpfungen ermitteln ließen, lag die Rate der dokumentbezogenen Verknüpfungen dort prozentual ebenfalls niedrig.

Diese Ergebnisse unterstreichen, dass ein DOI-Abgleich allein nicht ausreichen würde, um eine umfassende Anreicherung insbesondere der FIS Bildung Literaturdatenbank zu erreichen, die (von Sonderregelungen abgesehen) Literatur ab 1980 nachweist. Mit Blick auf den Sammlungszeitraum lässt sich generell sagen, dass ein DOI-basierter Abgleich nicht abdeckend sein kann, da der DOI – global gesehen – erst ab ca. Mitte der Nullerjahre in die weitere Verbreitung kam.⁸ Aus diesen Gründen wurde der DOI-Abgleich im Projekt nur als Ergänzung zur im folgenden Abschnitt beschriebenen Anreicherung über Werkbündelung im Culturegraph-Bestand der DNB genutzt.

4. Culturegraph-Werkbündelung und ihre Anwendung auf die Bestände des Fachportals Pädagogik

4.1 Culturegraph-Werkbündelung

Die DNB betreibt den Datenpool Culturegraph, in dem die Metadaten der deutschen Bibliotheksverbünde und die des Österreichischen Bibliothekenverbundes, ergänzt um die DNB-Bestände, aggregiert werden. Insgesamt umfasst Culturegraph so ca. 245 Millionen Datensätze und wird täglich über

8 What is a DOI?, <https://www.doi.org/the-identifier/what-is-a-doi/>, Stand: 18.09.2025.

Updates aus dem Gemeinsamen Verbündeindex (GVI) aktualisiert.⁹ Auf Basis des Culturegraph-Bestandes werden verschiedene Datenabgleichverfahren entwickelt und zum Teil auch bereits produktiv eingesetzt, die es ermöglichen, Informationen zur Anreicherung von Titeldaten zu ermitteln. Hierzu zählt die Werkbündelung, bei der zu einem Werk gehörende Publikationen in Bündeln zusammengeführt werden.¹⁰ Dies geschieht, indem aus verschiedenen Metadatenelementen charakteristische Schlüssel für jede Publikation erstellt werden. Es gibt zwei verschiedene Arten von Schlüsseln: Einerseits werden Systemnummern wie die ISBN, die Identifikationsnummer der erstkatalogisierenden Bibliothek (EKI) oder die „OCLC control number“ in Kombination mit dem Haupttitel verwendet; andererseits kommen auch Kombinationen aus erstellenden und beteiligten Personen und Körperschaften und vollständigen Titeln sowie alternativen und abweichenden Titeln zur Anwendung.¹¹

Die Schlüssel werden anschließend mit einem Suchalgorithmus abgeglichen. Sobald zwei Publikationen mindestens einen identischen Schlüssel aufweisen, werden sie zu einem Bündel zusammengefasst. Auf diese Weise werden auch verschiedene Auflagen, Ausgaben und Übersetzungen eines Werkes zusammengeführt und ein breit ausgelegter Werksbegriff verfolgt. Beispielhaft ist dies in Abb. 2 zu sehen, wo deutsche Übersetzungen eines Titels über den Einheitstitel mit originalsprachlichen Ausgaben in einem Bündel verknüpft werden, obwohl die jeweiligen Titel nicht in allen Ausgaben vorhanden sind. Jeweils die fettgedruckten bzw. kursiv fettgedruckten, übereinstimmenden Schlüssel sorgen für die Zusammenstellung des Bündels. Diese Werkbündel können dafür verwendet werden, werkbezogene Informationen unter Bündelmitgliedern auszutauschen.

⁹ Gemeinsamer Verbündeindex, <https://www.agv-gvi.de/>, Stand: 18.09.2025.

¹⁰ Die Werkbündelung in der DNB greift frühere Ansätze zur Bündelung von Publikationen auf, siehe z. B. Wiesenmüller, Heidrun; Pfeffer, Magnus: Abgleichen, anreichern, verknüpfen. Das Clustering-Verfahren – eine neue Möglichkeit für die Analyse und Verbesserung von Katalogdaten, in: BuB 65 (9), 2013, S. 625–629, https://www.b-u-b.de/fileadmin/archiv/imports/pdf_files/2013/bub_2013_09_625_629.pdf, Stand: 18.09.2025; Pfeffer, Magnus: Using Clustering Across Union Catalogues to Enrich Entries with Indexing Information, in: Spiliopoulou, Myra; Schmidt-Thieme, Lars; Janning, Ruth (Hg.): Data Analysis, Machine Learning and Knowledge Discovery, Cham 2014, S. 437–445. Und im Kontext des WorldCat: Gatenby, Janifer; Greene, Richard O.; Oskins, W. Michael; Thornburg, Gail: GLIMIR. Manifestation and Content Clustering within WorldCat, in: code{4}lib Journal 17, 2012, <https://journal.code4lib.org/articles/6812>, Stand: 18.09.2025; Gatenby, Janifer; Thornburg, Gail; Weitz, Jay: Collected Work Clustering in WorldCat, in: code{4}lib Journal 30, 2015, <https://journal.code4lib.org/articles/10963>, Stand: 18.09.2025.

¹¹ Vgl. Vorndran, Angela: Hervorholen, was in unseren Daten steckt! Mehrwerte durch Analysen großer Bibliotheksdatenbestände, in: o-bib. Das offene Bibliotheksjournal 5 (4), 2018, S. 166–180, <https://doi.org/10.5282/o-bib/2018h4s166-180>; Vorndran, Angela; Grund, Stefan: Metadata Sharing. How to Transfer Metadata Information among Work Cluster Members, in: Cataloging & Classification Quarterly 59 (8), 2021, S. 757–774, <https://doi.org/10.1080/01639374.2021.1989101>.



Abb. 2: Beispiel eines Werkbündels mit Publikationen in verschiedenen Sprachen, Quelle: Eigene Darstellung

Insbesondere mit einer Publikation verbundene Personen wie Autor*innen oder andere beteiligte Personen können von Informationen aus der Werkbündelung profitieren. Ist in dem Werkbündel bei einem Bündelmitglied ein GND-Identifer für einen Namen enthalten, kann dieser auf namensgleiche Personen anderer Bündelmitglieder übertragen werden. Existieren verschiedene GND-Identifer für einen Namen in einem Bündel, wird keiner übernommen, zusätzlich greifen weitere Plausibilitätskontrollen.¹² Somit wird eine größere Normierung und Vernetzung der Daten erreicht. Seit mehreren Jahren ermittelt die DNB so GND-Verknüpfungen zur Anreicherung in ihren Daten. Seit 2021 wurden zunächst in Retrobearbeitungen des Altbestandes und seit Ende 2022 in einem täglichen Verfahren mittlerweile über 3 Millionen Personenverknüpfungen ergänzt.

Die Verknüpfung zur GND ermöglicht eine eindeutige Identifikation einer Person, unterstützt die Vernetzung der Daten und erleichtert die Bereitstellung weiterer personenbezogener Informationen. Seit 2021 besteht auch das Angebot für die an Culturegraph beteiligten Verbünde und externe Interessierte, Vorschläge für Personenverknüpfungen für Verbunddaten sowie eine aggregierte Datei der Werkbündelungsergebnisse zu beziehen.¹³

Das hier vorgestellte Projekt geht darüber hinaus, indem es ermöglicht, externe Datenbestände von Institutionen, die nicht an einem Bibliotheksverbund beteiligt sind, punktuell in die etablierten Workflows der Werkbündelung und Ermittlung von Personenverknüpfungen zu integrieren. Liegen die bereitgestellten Daten in einem passenden Format vor, können sie nach einer initialen inhaltlichen

12 Vgl. Glagla-Dietz, Stephanie; Grund-Davidov, Stefan; Habermann, Nicole; Vorndran, Angela: Personen in GND und Titeldaten. Vorschläge, Anreicherungen und Verknüpfungen, in: o-bib. Das offene Bibliotheksjournal 11 (1), 2024, S. 1–17, <https://doi.org/10.5282/o-bib/5998>.

13 Monatlich aktualisierte Versionen der Dateien können unter <https://data.dnb.de/culturegraph/> bezogen werden.

Prüfung nach denselben Konfigurationen weiterverarbeitet werden wie die bereits in Culturegraph enthaltenen Daten.

4.2 Datenaustausch

Der Datenaustausch zwischen DNB und den an Culturegraph beteiligten Verbundsystemen basiert auf dem MARC 21-Standard. Ebenso nutzt auch das DIPF den MARC 21-Standard für die Bereitstellung von Datenabzügen der FIS Bildung. Dementsprechend konnte für die Kooperation beiderseits auf bestehende Strukturen zurückgegriffen werden. Das Resultat des Verfahrens, im Wesentlichen eine Liste von lokalen Dokument-Identifiern mit den zugehörigen Personen-GND-Identifiern, wurde als CSV-Datei bereitgestellt.

4.3 Einbinden in Werkabgleich

Im Rahmen dieses Projekts wurden für die vom DIPF bereitgestellten Titeldaten Schlüssel nach den oben beschriebenen Mustern erzeugt. Diese wurden mit den aktuellen Schlüsseln der Datensätze in Culturegraph abgeglichen und über den Suchalgorithmus zu neuen Bündeln unter Beteiligung der FID-Datensätze zusammengeführt. Analog zur oben beschriebenen Vorgehensweise wurden vorhandene GND-Verknüpfungen für Personen in FID-Datensätzen von namensgleichen Personen in Werkbündeln ermittelt.

4.4 Methode zur Prüfung der Ergebnisse

Zur Bewertung der vom Werkbündelungsverfahren ausgegebenen GND-Identifizier für Personen wurde pro Datenquelle (FIS Bildung, ERIC) zunächst eine quantitative Auswertung vorgenommen. Ziel war es, die Rücklaufzahlen des Verfahrens zu analysieren und insbesondere nach Dokumenttypen (Monografie, Zeitschriftenaufsatz, Sammelwerksbeitrag) zu differenzieren. Diese Untergliederung diente der Plausibilitätskontrolle, ob alle gelieferten Dokumente verarbeitet wurden und um einen Eindruck von der Größenordnung der Anreicherungen zu bekommen.

Da es sich bei diesem Anreicherungsverfahren um einen automatisierten und unbeaufsichtigten Prozess handelt und die Ergebnisse nicht einzeln geprüft werden können, kommt der Abschätzung der Korrektheit von gelieferten Verknüpfungen besondere Bedeutung zu. Als weiterer zentraler Aspekt ist die Vollständigkeit wichtig, d. h. ob das Verfahren die optimale Verknüpfungsanzahl im Vergleich zu einem intellektuellen Verfahren liefert. Eine vollständige Abdeckung ist zwar anzustreben, aber praktisch schwer zu erreichen. Daher stand die Prüfung der Korrektheit oder Fehlerrate im Vordergrund, zumal die Prüfung der Vollständigkeit mit einer zeitintensiven Recherche in der GND einhergegangen wäre. Die Bewertung erfolgte auf Basis manuell ausgewerteter Stichproben.

A		B		C		D	
1	Fid	Fachportal_Doktyp		DNB Abgleich_gefundene_Namen		DNB Abgleich_gefundene_GND_IDs	
2	3376721	Monografie		Blank-Sangmeister, Ursula;Schüffel, Janet		108997448;121505727X	
3	3317066	Monografie		Heidkamp, Birte;Kergel, David		119937802X;109324724X	
4	3317143	Monografie		Becker-Mrotzek, Michael;Geyer, Sabrina;Hasselhorn, Marc		121290468;1156329973;122500857;1013943678;1031117199;132700638;129974218;115633022X	
5	3377998	Monografie		Klieme, Eckhard;Köller, Olaf;Weis, Mirjam		112732283;118137093;1082145238	
6	3378582	Sammelwerksbeitrag		Brefeld, Ulf;Schreiber, Waltraud;Trautwein, Ulrich;Wagner		135704669;121941027;126261598;133960455	
7	3345566	Monografie		Keil, Andreas;Kuckuck, Miriam		121913783;1058656619	
8	3379412	Monografie		Finn, Kirsty;Gagnon, Jessica;Gray, Emily		107707753X;1203102844;1033825689	
9	3266436	Monografie		Hoffmann, Ludger;Kameyama, Shin'ichi;Sahiner, Pembe;W		132053225;128858524;1156791235;1139780573	
10	3346505	Monografie		Gaile, Dorothee;Kaiser, Carl;Müller-Platt, Anne;Schmidt-D		130930474;123777425X;1210561336;1210561557;1210561670	

Abb. 3: Stichprobe aus Werkbündelungsverfahren, zur Auswertung tabellarisch aufbereitet, Quelle: Eigene Darstellung

Dabei wurden einzelne angereicherte Literaturnachweise zufällig ausgewählt (vgl. Abb. 3) und die hergestellten Verknüpfungen mit Informationen aus den zugehörigen GND-Sätzen abgeglichen (s. Tab. 2). Eine Verknüpfung wurde als korrekt eingestuft, wenn sich aus dem GND-Satz eine Beteiligung an der jeweiligen Publikation entweder durch eine explizite Werkverknüpfung oder durch das Gesamtbild der biografischen und bibliografischen Angaben ableiten ließ.

Tab. 2: Auswertung des Werkbündelungsverfahrens: Prüfergebnisse der Stichprobe

E	F	G	H	I	J
Anz_Personen_Fachportal	Anz_ermittelbare_GND-IDs	Anz_geliefert_DNB	Anz_korrekte_DNB	P_DNB	R_DNB
4	3	2	2	1	0,667
2	2	2	2	1	1
9	9	8	8	1	0,899
4	4	3	3	1	0,75
4	4	4	4	1	1
3	3	2	2	1	0,667
5	5	3	3	1	0,6
5	5	4	4	1	0,8
5	5	5	5	1	1

Die Korrektheit wurde anhand der Trefferquote bewertet, indem die Anzahl der korrekt gelieferten Verknüpfungen mit der Gesamtanzahl der gelieferten Verknüpfungen in Beziehung gesetzt wurde (s. Tab. 2: Spalten H und G). Die Bewertung der Vollständigkeit orientiert sich methodisch an gängigen Definitionen der Kenngröße „Recall“ im Information Retrieval.¹⁴ Zur Berechnung wurde die Anzahl der korrekt gelieferten Verknüpfungen in Beziehung gesetzt zur ermittelbaren Anzahl der Verknüpfungen (s. Tab. 2: Spalten H und F). Diese Anzahl war im Projekt nicht von vornherein bekannt und musste intellektuell durch Recherche in der GND ermittelt werden. Werte für Korrektheit und Vollständigkeit wurden zunächst zeilenweise pro Publikation bestimmt und anschließend zu einer Gesamtaus-sage zusammengefasst (s. Tab. 2: Spalten I und J). Die Kenngrößen sind in Anlehnung an „Precision“ (Genauigkeit) und „Recall“ (Vollständigkeit) mit P_DNB und R_DNB beschriftet.

14 Petras, Vivien; Womser-Hacker, Christa: Evaluation im Information Retrieval, in: Kuhlen, Rainer; Lewandowski, Dirk; Semar, Wolfgang u. a. (Hg.): Grundlagen der Informationswissenschaft, 7., völlig neu gefasste Ausgabe, Berlin; Boston 2023, S. 443–453, <https://doi.org/10.1515/9783110769043-038>.

4.5 Ergebnisse

FIS Bildung

Für die FIS Bildung Literaturdatenbank konnte mittels Werkbündelungsverfahren eine große Anzahl an Verknüpfungen zwischen Personen in Titeldaten und GND-Identifiern erzeugt werden. Tabelle 3 zeigt die Resultate nach Dokumenttyp differenziert:

Tab. 3: Resultate des Abgleichverfahrens der Werkbündelung für FIS Bildung im Überblick

Dokumenttyp	Anzahl Dokumente in Abgleich	Anzahl Dokumente mit vorgeschlagenen Anreicherungen	Anzahl anzu-reichernde Verknüpfungen
Monografie	344.326	153.499 (44,6 %)	216.132
Sammelwerksbeitrag	126.338	7.122 (5,6 %)	9.083
Zeitschriftenaufsatz	659.829	88.031 (13,3 %)	111.600
Σ	1.130.493	248.652 (21,1 %)	336.815

Die aggregierten Zahlen zeigen, dass insbesondere beim Dokumenttyp Monografie eine hohe Verknüpfungsrate erzielt wurde. Das Verfahren war weniger ertragreich bei Sammelwerksbeiträgen und Zeitschriftenaufsätzen, was sich vermutlich mit der vorhandenen Datengrundlage in den Culturegraph-Daten erklären lässt.

Zur Qualitätsprüfung der gelieferten Daten wurde eine erste Stichprobe aus jenen Literaturnachweisen gezogen, für die im Fachportal bereits manuell vergebene Verknüpfungen vorlagen. In dieser Stichprobe waren 225 angereicherte Literaturnachweise enthalten, davon 130 Monografien und 95 Sammelwerksbeiträge. Insgesamt waren 481 Personen-Publikation-Verknüpfungen zu überprüfen, von denen 475 als korrekt eingestuft wurden. Daraus ergibt sich eine Genauigkeit von ca. 98,7 %.

Ergänzend wurde die Vollständigkeit (Recall) anhand derselben Stichprobe abgeschätzt. Die manuelle Recherche in der GND ergab, dass noch für weitere Personennamen ein GND-Identifizier ermittelt werden konnte. Das Verfahren hätte bestenfalls, die recherchierbaren Zuordnungen eingerechnet, 583 Verknüpfungen herstellen können. Die tatsächlich korrekt gelieferten 475 Verknüpfungen entsprechen damit einer Recall-Rate von rund 81,5 %. Hierbei ist allerdings nochmals hervorzuheben, dass die Stichprobe auf solche Literaturnachweise zurückgriff, für die auch in FIS Bildung bereits teilweise GND-Verknüpfungen bekannt waren; insofern ist die Vollständigkeit bei einer komplett zufälligen Ziehung einer Stichprobe aus dem Gesamtbestand der FIS Bildung möglicherweise niedriger als hier ermittelt.

Eine zweite Stichprobe untersuchte die Robustheit des Verfahrens gegenüber potenziellen Verwechslungen bei häufigen Nachnamen. Die Auswahl umfasste 50 Literaturnachweise vom Typ Monografie, in denen mindestens eine Person mit häufigem Nachnamen beteiligt war.¹⁵ Alle 90 Verknüpfungen wurden als korrekt bewertet.

Die Prüfung von Zeitschriftenaufsätzen erfolgte in einer weiteren Stichprobe, wobei jedoch auf die zeitintensive Ermittlung der Vollständigkeit verzichtet wurde. Es wurden 150 angereicherte Zeitschriftenaufsätze überprüft, wobei 267 Verknüpfungen kontrolliert wurden.¹⁶ Bei der Prüfung trat vermehrt die Situation auf, dass insbesondere bei Personen aus dem englischsprachigen Raum nur „rumpfhafte“ GND-Sätze vorhanden waren. Zwar passte beispielsweise öfter der angegebene Wirkungszeitraum zum Veröffentlichungsjahr der Publikation. Es war aber dennoch nicht eindeutig entscheidbar, ob es sich um eine korrekte Verknüpfung handelte. Die Bewertung der Verknüpfungen wurde deshalb auf „richtig“ oder „unklar“ umgestellt. Es wurden 239 Verknüpfungen oder 89,5 % als „richtig“ gewertet und 28 Fälle bzw. 10,5 % als „unklar“, wobei sich unter den unklaren vermutlich einige korrekte Verknüpfungen befinden.

Die Ergebnisse führen zu der Einschätzung, dass die Abgleichergebnisse der FIS Bildung Literaturdatenbank die Qualitätsanforderungen an Verknüpfungen, wie sie für die Personenprofilseiten erforderlich sind, erfüllen. Die Auswertung belegt die hohe Genauigkeit des Werkbündelungsverfahrens bei Monografien und Sammelwerksbeiträgen. Eine Verwechslung von Personen mit häufigen Nachnamen hat sich nicht bestätigt, könnte aber tiefer untersucht werden. Die Genauigkeit bei Zeitschriftenaufsätzen ist ebenfalls gut, wobei die Abdeckung – in Bezug auf die insgesamt angereicherten Nachweise sowie auf Nachweisebene für einzelne Verknüpfungen – aufgrund der Datenlage in Culturegraph erwartbar niedriger ausfällt.

ERIC

Für die ERIC-Datenbank, eine umfangreiche internationale Sammlung bildungs- und erziehungswissenschaftlicher Literatur, wurde das Werkbündelungsverfahren ebenfalls angewendet. Die Datenbank ist wie FIS Bildung in das Fachportal Pädagogik integriert und wird häufig ergänzend zur Recherche genutzt. Für die bestandsübergreifende Suche sind Schlagwortkonkordanzen zwischen dem FIS Bildung Schlagwortbestand und ERIC-Schlagwörtern implementiert. Der ERIC-Bestand gliedert sich in zwei Hauptdokumenttypen: „EJ“ für Zeitschriftenaufsätze und „ED“ für sonstige Materialien wie Monografien, Sammelwerksbeiträge, Berichte oder Rechtsnormen („other documents“). Eine Publikation in englischer Sprache ist notwendige Voraussetzung für deren Nachweis in ERIC.

Tabelle 4 zeigt die Ergebnisse des Abgleichs. Im Vergleich zu FIS Bildung fällt die Anreicherungsquote sehr viel geringer aus. Dies war erwartbar, da ERIC ausschließlich englischsprachige Literatur enthält und einen hohen Anteil an internationalen Zeitschriftenaufsätzen sowie Grauer Literatur aufweist. Es ist plausibel anzunehmen, dass diese Arten von Inhalten im Datenkorpus von Culturegraph wenig repräsentiert sind. Angesichts des insgesamt geringen Rücklaufs und des internationalen Bezugsrahmens wurde auf eine Prüfung der Vollständigkeit wegen der aufwändigen Recherchearbeit verzichtet.

15 Die Liste von häufigen Nachnamen wurde zusammengestellt auf Grundlage von „Lists of most common surnames in European countries“, https://en.wikipedia.org/wiki/Lists_of_most_common_surnames_in_European_countries, Stand: 18.09.2025.

16 Die auf Vortragsfolie 10 abgebildete Zahl zu diesem Punkt ist nicht korrekt. Dort sind 267 Zeitschriftenaufsätze genannt, richtig sind 150. Siehe <https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/19789>, Stand: 18.09.2025.

Zur Qualitätsprüfung der gelieferten Verknüpfungen wurde eine Stichprobe von 399 angereicherten Literaturnachweisen gezogen. Sie bestand aus 199 Zeitschriftenaufsätzen (EJ) und 200 sonstigen Dokumenten (ED) und deckte die Erscheinungsjahre von 1964 bis 2022 ab. Insgesamt wurden 495 Verknüpfungen überprüft. Davon entfielen 264 Verknüpfungen auf EJ und 231 auf die Kategorie ED. Die Bewertung erfolgte analog zur FIS-Stichprobe bei Zeitschriftenaufsätzen durch Abgleich mit GND-Sätzen.

Insgesamt wurden 416 Verknüpfungen (ca. 84 %) als korrekt eingestuft. In 6 Fällen (1,2 %) wurde eine fehlerhafte Verknüpfung identifiziert, 83 Verknüpfungen (16,8 %) konnten nicht eindeutig beurteilt werden. Die unklaren Fälle waren erneut darauf zurückzuführen, dass die zugehörigen GND-Sätze nur sehr knappe Informationen enthielten. In der Kategorie EJ wurde eine leicht bessere Genauigkeit gegenüber ED beobachtet.

Tab. 4: Resultate des Abgleichverfahrens für ERIC im Überblick

Dokumenttyp	Anzahl Dokumente in Abgleich	Anzahl Dokumente mit vorgeschlagenen Anreicherungen	Anzahl anzureichernde Verknüpfungen
EJ (Zeitschriftenaufsatz)	1.305.303	13.596 (1 %)	18.572
ED („Andere“)	604.924	8.868 (1,5 %)	10.536
Σ	1.910.227	22.464 (1,2 %)	29.108

Die Ergebnisse zeigen, dass das Werkbündelungsverfahren hinsichtlich Genauigkeit auch für internationale Bestände grundsätzlich tauglich ist. Die erreichbare Abdeckung bleibt allerdings hinter stärker „GND-nahen“ Beständen, wie der FIS Bildung Literaturdatenbank, zurück. Die festgestellte Genauigkeit ist dennoch ausreichend, um die gewonnenen Verknüpfungen im Projekt weiter zu verwenden.

5. GND-ORCID-Konkordanzen

Gerade aus Sicht des FID ist die ORCID iD von besonderem Interesse, da es sich um ein in der Wissenschaft genutztes System zur eindeutigen Identifikation von Personen handelt, das zudem von Wissenschaftler*innen selbst inhaltlich gepflegt wird. ORCID-Profile enthalten häufig weiterführende und aktuellere Informationen als die GND, z. B. in Bezug auf Namensänderungen, akademische Grade, die aktuelle Affiliation, bis hin zu selbst verfassten Lebensläufen. Diese Daten stehen als strukturierter Open-Data-Abzug zur Verfügung und können zudem einzeln über die jeweilige ORCID-URI maschinenlesbar abgerufen werden.¹⁷ Vor diesem Hintergrund ist die Verwendung von ORCID iDs für die geplanten Personenprofilseiten besonders attraktiv.

¹⁷ Bei ORCID können Personendaten in diversen technischen Formaten angefordert werden, siehe https://github.com/ORCID/ORCID-Source/blob/main/CONTENT_NEGOTIATION.md, Stand: 18.09.2025.

In über 300.000 Personendatensätzen der GND sind bereits ORCID-Identifizierer verknüpft. Diese Datengrundlage konnte genutzt werden, um die DOI-ORCID-Abgleiche mit GND-Identifiern zu erweitern und umgekehrt aus den GND-Verknüpfungen des Werkbündelungsverfahrens die ORCID-Identifizierer der Personen zu ergänzen.

Die ERIC-Datenbank als Beispiel enthält bereits über 80.000 verschiedene ORCID-Identifizierer mit insgesamt mehr als 130.000 Dokumentverknüpfungen. Durch Anwendung der Konkordanz ließen sich 6.319 GND-Identifizierer für Personen zuordnen, die an 12.576 Publikationen beteiligt waren. Beim „GND-nahen“ Bestand FIS Bildung Literaturdatenbank waren im Crossref-Abgleich 4.700 ORCID iDs ermittelt worden, für die es in 1.298 Fällen eine GND-Konkordanz gab (s. Tab. 1).

Allerdings ergeben sich aus der Kombination von GND und ORCID auch Herausforderungen. Während ein GND-Satz redaktionell gepflegt und stabil ist, kann ein ORCID-Profil jederzeit bearbeitet oder gelöscht werden. Zudem variiert die Informationsdichte der ORCID-Profile. Viele enthalten nur minimale Angaben, was ihre Nutzbarkeit als Quelle von biografischen Informationen für Personenprofile einschränkt; dies trifft jedoch auch teilweise auf GND-Sätze zu. Eine weitere offene Frage betrifft die technische und gestalterische Integration beider Identifizierer-Systeme in ein konsistentes Modell: Welcher Identifizierer sollte im Profil „führend“ sein? Wie lässt sich das dynamische Entstehen oder Wegfallen von GND-ORCID-Konkordanzen (und damit neuen Dokumentverknüpfungen) sinnvoll auffangen? Welche Regeln gelten für die Zusammenführung biografischer Angaben aus zwei Quellen, die sich potenziell auch widersprechen können? Für den Aufbau der Personenprofile im Fachportal stellt dies eine zentrale konzeptionelle Aufgabe dar, die sowohl technische Anforderungen als auch die Bedürfnisse und Erwartungen der wissenschaftlichen Zielgruppen berücksichtigen muss.

6. Fazit und Ausblick

Die geringe Fehlerrate des Anreicherungsverfahrens ist aus Sicht des FID überzeugend, so dass die gewonnenen Verknüpfungen in die Datenbasis für Personenprofilseiten übernommen werden. Für die nun verknüpften GND- und ORCID-Identifizierer wurden die zugehörigen Personendatensätze in eine Datenbank integriert, in der die Zusammenführung von Personeninformationen beider Quellen technisch realisiert wird. Insbesondere die bei FIS Bildung erreichte Abdeckung ist wegen ihrer zentralen Stellung für den deutschsprachigen Raum von großer Bedeutung. Gemeinsam mit den ergänzenden Ansätzen besteht eine Perspektive, wie die Abdeckung von Personen-Identifiern in Beständen des Fachportals Pädagogik kontinuierlich erweitert werden kann. Geplant ist zudem eine Funktion im Fachportal, die es ermöglicht, Personenprofile mit Literaturdatensätzen zu verknüpfen, um die Fachcommunity aktiv in den Ausbau der Datenbasis einzubeziehen („Claiming“). Für Personen mit einem aktiv gepflegten ORCID-Profil bietet die Verknüpfung der Personenseite mit dem ORCID-Account den Vorteil, dass alle relevanten Informationen zentral in ORCID gepflegt werden können. Dadurch bleibt auch die Profilseite automatisch aktuell, ohne dass ein weiteres Profil separat gepflegt werden muss.

Für andere Informationseinrichtungen mit vergleichbarer Ausgangslage stellt das Verfahren der DNB eine interessante Möglichkeit dar, eigene Metadatenbestände systematisch mit GND-Normdaten anzureichern. In Portalen, die bereits GND-basierte Personenseiten für eine wissenschaftliche Zielgruppe anbieten, kann die ergänzende Integration von ORCID-Identifiern eine sinnvolle Erweiterung

darstellen. Der dateibasierte Austausch im MARC(XML)-Format ermöglicht eine einfache technische Anbindung, ohne dass aufwändige Maßnahmen zur Integration erforderlich sind.

Aus Sicht der Deutschen Nationalbibliothek ergibt sich aus dem mittlerweile mehrjährigen erfolgreichen Einsatz des Verfahrens der Werkbündelung zur Übertragung von Personenverknüpfungen im eigenen Bestand und als Angebot an Bibliotheksverbünde die Gelegenheit, diesen Service auch für Institutionen anzubieten, die nicht an einem Bibliotheksverbund beteiligt sind. Durch eine modulare Gestaltung des Workflows bei der Erstellung von Werkbündeln und der Ermittlung von fehlenden Personenverknüpfungen ist es möglich, auch externe Datenbestände in den Ablauf zu integrieren und für diese Personenverknüpfungen zu liefern.

Es besteht für interessierte Institutionen die Möglichkeit, ihre Daten für die Teilnahme an der Werkbündelung direkt an die DNB zu übermitteln, wenn sie im Format MARC 21 oder MARC XML vorliegen. Wie sich in Bezug auf die verschiedenen Datenbestände des Fachportals Pädagogik zeigt, hängt der quantitative Erfolg des Verfahrens auch von der Art der Bestände und der Übereinstimmung mit der Ausrichtung des in Culturegraph aggregierten Bestandes ab. Qualitativ wird in allen Fällen für die ermittelten Verknüpfungen ein einheitlich hohes Niveau erreicht.

Über den hier vorgestellten Anwendungsfall hinaus liefern Culturegraph-Werkbündel außerdem auch andere werkbezogene Metadatenelemente wie Klassifikationsnotationen (z. B. DDC, RVK) oder verbale Sacherschließung, die sich für eine Anreicherung von Metadaten eignen. Für die Ermittlung von Metadatenelementen, die sich nicht auf die Werkebene beziehen, sondern manifestationsbezogen beispielsweise nur für einzelne Auflagen einer Publikation relevant sind, bietet Culturegraph außerdem eine Manifestationsbündelung an. Auch diese ist mit einem analogen Workflow für Daten externer Institutionen nutzbar und kann hier Anreicherungsoptionen bieten.

Das Verfahren bietet somit in vielfältiger Weise Möglichkeiten, in deutschsprachigen Bibliotheken intellektuell erstellte Inhalte breit zu nutzen und kooperativ weiterzuverwenden. Im Sinne einer offenen Dateninfrastruktur können so bereits vorliegende, qualitativ hochwertige und verlässliche Daten einen umfangreichen Nutzen innerhalb und außerhalb bibliothekarischer Verbundsysteme finden. Eine Ausweitung der hier vorgestellten prototypischen Integration von externen Daten in Culturegraph-Prozesse ist sowohl in Hinsicht auf weitere Domänen über die Bildungsforschung hinaus, als auch mit dem Ziel der Anreicherung weiterer Inhalte, z. B. inhaltserschließender Elemente, erfolgversprechend.

Literaturverzeichnis

- Gatenby, Janifer; Greene, Richard O.; Oskins, W. Michael; Thornburg, Gail: GLIMIR. Manifestation and Content Clustering within WorldCat, in: code{4}lib Journal 17, 2012, <https://journal.code4lib.org/articles/6812>, 18.09.2025.
- Gatenby, Janifer; Thornburg, Gail; Weitz, Jay: Collected Work Clustering in WorldCat, in: code{4}lib Journal 30, 2015, <https://journal.code4lib.org/articles/10963>, Stand: 18.09.2025.
- Glagla-Dietz, Stephanie; Grund-Davidov, Stefan; Habermann, Nicole; Vorndran, Angela: Personen in GND und Titeldaten. Vorschläge, Anreicherungen und Verknüpfungen, in: o-bib. Das offene Bibliotheksjournal 11 (1), 2024, S. 1–17, <https://doi.org/10.5282/o-bib/5998>.

- Jungmann, Marcel; Vorndran, Angela: Normdatenverknüpfungen aus Culturegraph für die Bildungsforschung (und darüber hinaus), Vortrag auf dem 9. Bibliothekskongress (zugleich 113. BiblioCon) in Bremen, 26.06.2025, Vortragsfolien: <https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/19789>, Stand: 18.09.2025.
- Meadows, Alice; Haak, Laurel L.; Brown, Josh: Persistent Identifiers. The Building Blocks of the Research Information Infrastructure, in: Insights 32, 2019, <https://doi.org/10.1629/uksg.457>.
- Oerder, Thomas; Schindler, Christoph; Rittberger, Marc: Vom Fachinformationssystem Bildung zum Fachportal Pädagogik. Die Gestaltung der Digitalisierung der Wissenschaft am Beispiel der Bildungsforschung, in: Bibliotheksdienst 52 (3/4), 2018, S. 196–207, <https://doi.org/10.1515/bd-2018-0027>.
- Petras, Vivien; Womser-Hacker, Christa: Evaluation im Information Retrieval, in: Kuhlen, Rainer; Lewandowski, Dirk; Semar, Wolfgang u. a. (Hg.): Grundlagen der Informationswissenschaft, 7., völlig neu gefasste Ausgabe, Berlin; Boston 2023, S. 443–453, <https://doi.org/10.1515/9783110769043-038>.
- Pfeffer, Magnus: Using Clustering Across Union Catalogues to Enrich Entries with Indexing Information, in: Spiliopoulou, Myra; Schmidt-Thieme, Lars; Janning, Ruth (Hg.): Data Analysis, Machine Learning and Knowledge Discovery, Cham 2014, S. 437–445.
- Vorndran, Angela: Hervorholen, was in unseren Daten steckt! Mehrwerte durch Analysen großer Bibliotheksdatenbestände, in: o-bib. Das offene Bibliotheksjournal 5 (4), 2018, S. 166–180, <https://doi.org/10.5282/o-bib/2018h4s166-180>.
- Vorndran, Angela; Grund, Stefan: Metadata Sharing. How to Transfer Metadata Information among Work Cluster Members, in: Cataloging & Classification Quarterly 59 (8), 2021, S. 757–774, <https://doi.org/10.1080/01639374.2021.1989101>.
- Wiesenmüller, Heidrun; Pfeffer, Magnus: Abgleichen, anreichern, verknüpfen. Das Clustering-Verfahren – eine neue Möglichkeit für die Analyse und Verbesserung von Katalogdaten, in: BuB 65 (9), 2013, S. 625–629, https://www.b-u-b.de/fileadmin/archiv/imports/pdf_files/2013/bub_2013_09_625_629.pdf, Stand: 18.09.2025.