CC BY

Datenbereinigung in GOKB und ZDB

Christin Seegert

Verbundzentrale des GBV (VZG)

GOKB

Andrea Stei

Staatsbibliothek zu Berlin

ZDB

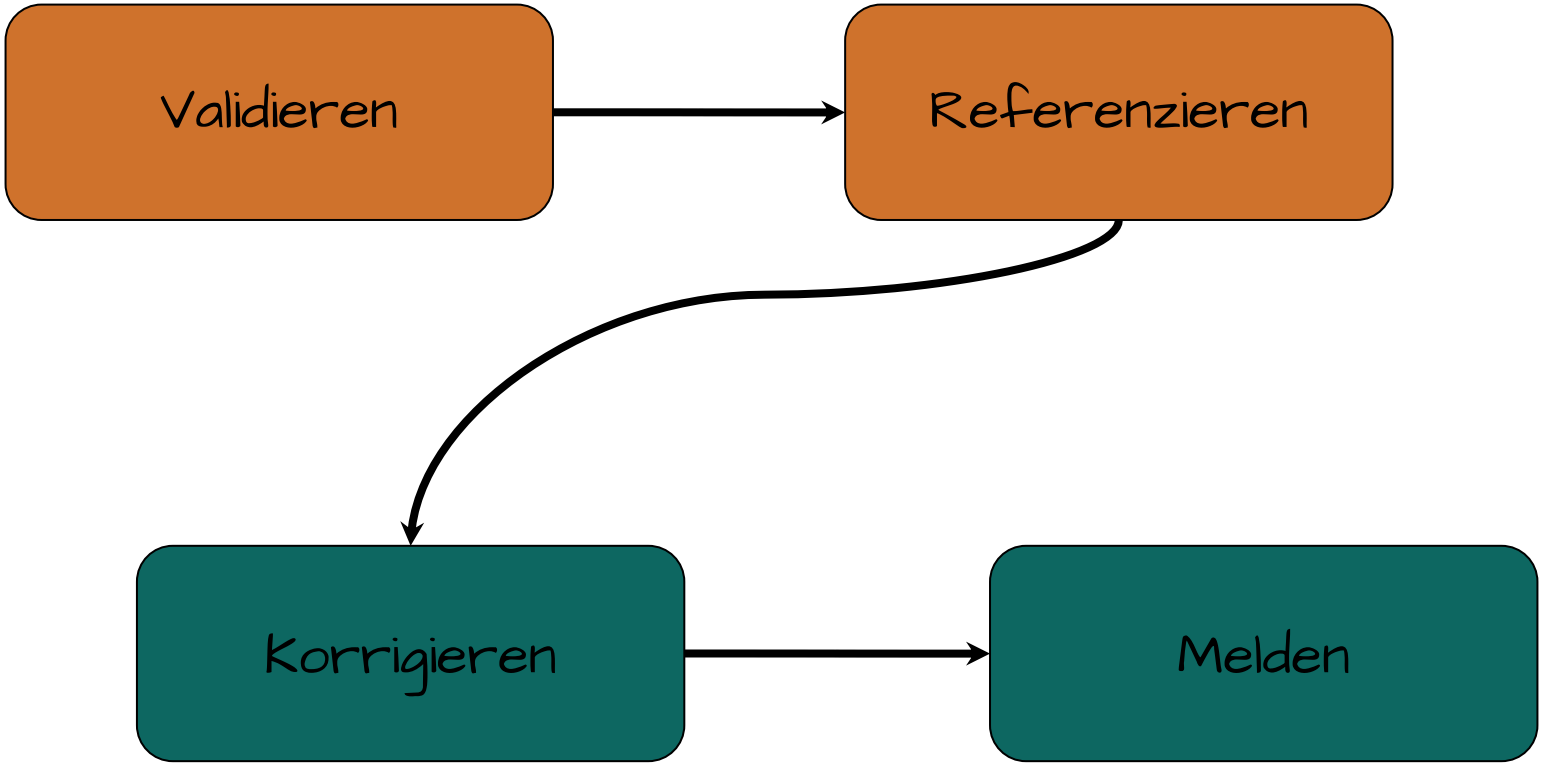
Die GOKB ist…

- … eine **communitybasierte globale Austauschplattform**
  - für **Metadaten** elektronischer Bücher, Zeitschriften und anderer Medienarten
  - die in **Paketen** (Anbieterpakete, konsortiale und lokale Pakete) organisiert sind
  - um **Prozesse** in nachgelagerten Systemen (z.B. zur **Erwerbung**) zu unterstützen

Die GOKB liefert Daten, die…

- eindeutig **referenziert** sind.
- **angereichert** wurden (Identifikatoren, Namensansetzung, Erscheinungshistorie).
- **bereinigt, eindeutig** und **korrigiert** sind.

Datenbereinigung in der GOKB


Validieren
Referenzieren
Korrigieren
Melden

Validierung

Daten vorher prüfen: Unser KBART-Validator

Ergebnisse

Dateistruktur - Warnungen
- Fehlende Spalte: notes
- Fehlende Spalte: publisher_name

Titellisten, die Warnungen erzeugen, können in die GOKB importiert werden, auch wenn es Auffälligkeiten gibt. Es wird dringend empfohlen, vorab die Auffälligkeiten auf ihre Unbedenklichkeit zu prüfen.

Zeilenstatistik
Insgesamt: 3 Warnungen: 0 Fehler: 1

Spaltenstatistik - Fehler
- **online_identifier** - 1

Es wurden Fehler in einzelnen Titeldaten festgestellt. Die Datei kann erst importiert werden, wenn diese Fehler behoben wurden.

Fehler

Warnungen

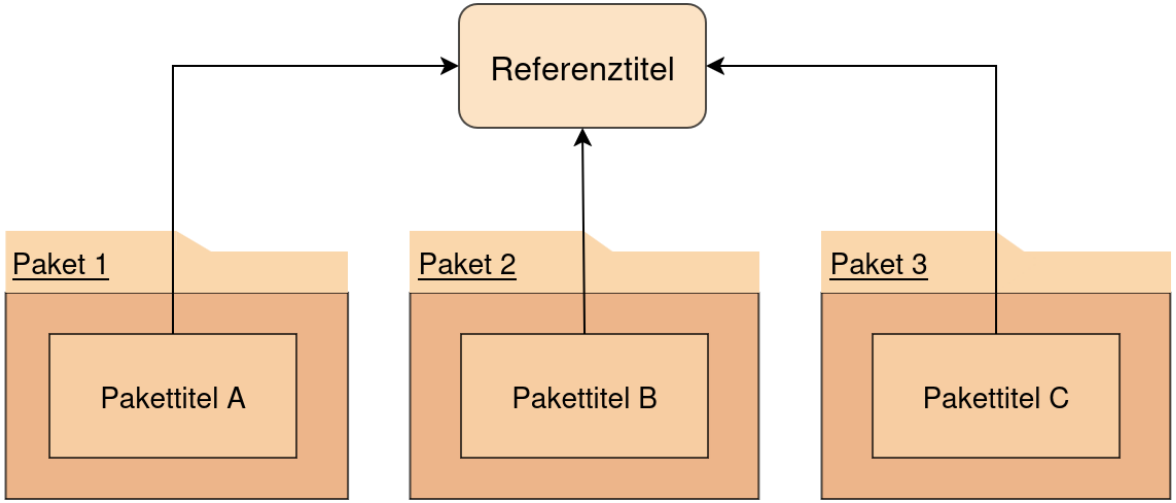
ERGEBNIS EXPORTIEREN

	A	B	C	D
1	Zeile	Spalte	Ursache	Typ
2	n.a.	notes	Fehlende Spalte:	Dateistruktur-Warnung
3	n.a.	publisher_name	Fehlende Spalte:	Dateistruktur-Warnung
4	1	online_identifier	Wert '0340-7900' ist fehlerhaft!	Fehler

Referenzierung

Referenzierung von Titeln

- Informationen aus verschiedenen Pakettiteln werden in einem “Mastertitel” gesammelt



Der Referenztitel

Name *

Journal of the Operational Research Society

NAME BEARBEITEN

Status

Aktuell | Archiviert | Erwartet | Löschen

UUID

0a4cca39-d4ee-4bfc-a30d-6911bc366b53

OA-Status: Unbekannt

Medientyp: Zeitschrift

Erscheinungsbeginn: 1978-01-01

Erscheinungsende

IDENTIFIKATOREN 7 | VERLAGE 2 | NAMENSVARIANTEN 0 | PRÜFANFRAGEN 0 | PAKETTITEL 13 | TITELVERLAUF 0 | SACHERSCHLIESSUNG 1

+ HINZUFÜGEN | ENTFERNEN

Namensraum	Wert
EZB-ID	18956
EZB-ID	2877
Taylor & Francis	tjor20
Taylor & Francis	joperresesoci
ZDB-ID	2007775-0
e-ISSN	1476-9360
p-ISSN	0160-5682

Der Referenztitel

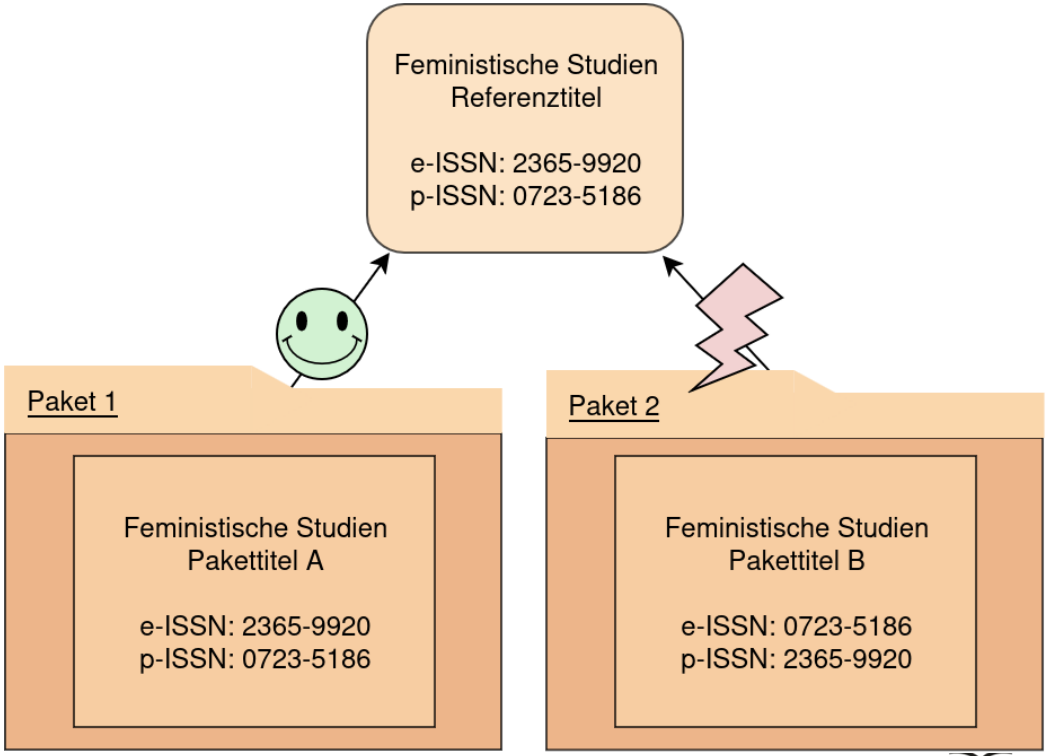
Namensraum	Wert
EZB-ID	18956
EZB-ID	2877
Taylor & Francis	tjor20
Taylor & Francis	joperresesoci
ZDB-ID	2007775-0
e-ISSN	1476-9360
p-ISSN	0160-5682

Der Referenztitel

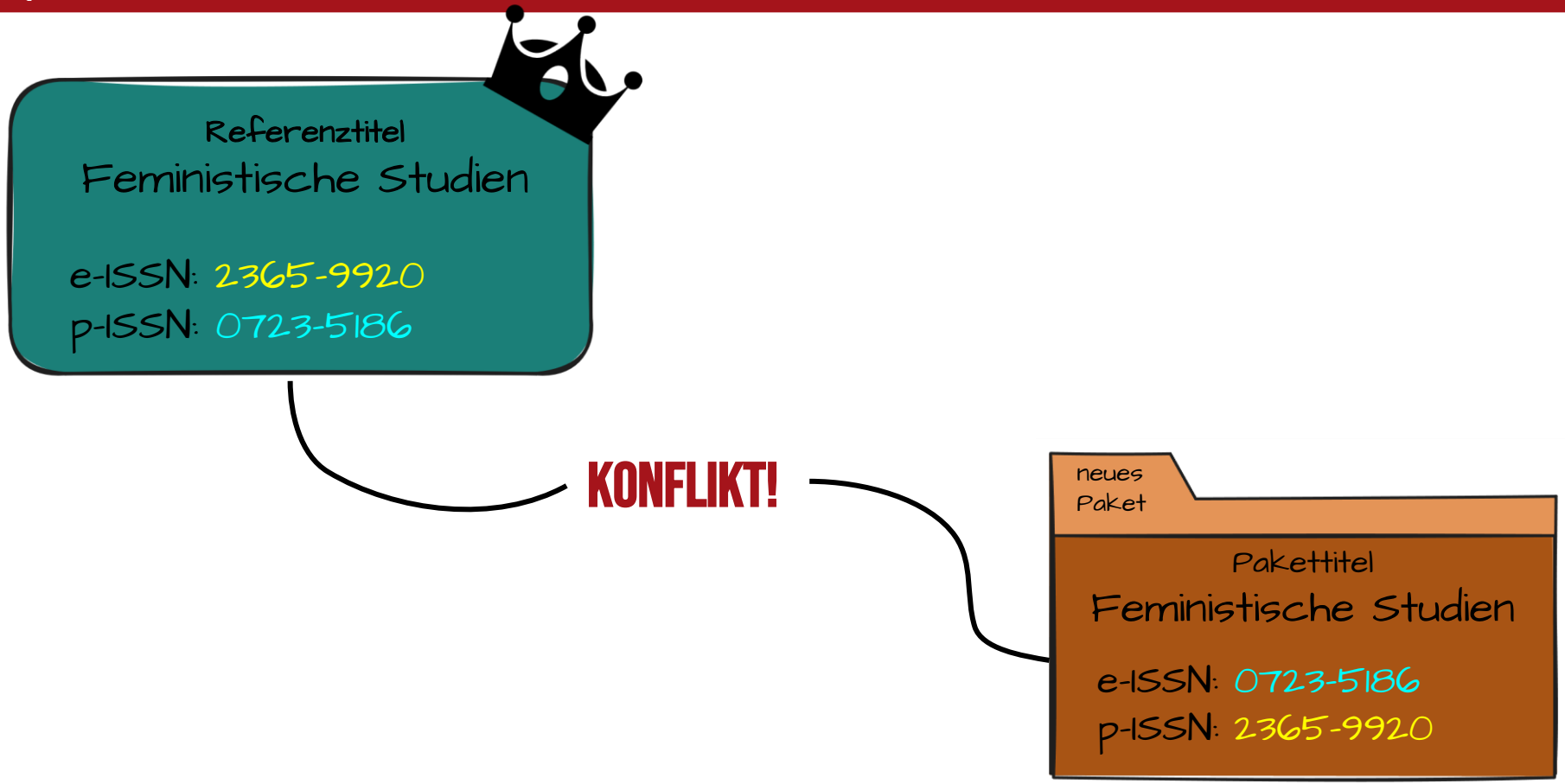
- bibliographisch aufbereiteter Titel
- dient in der GOKB als “Mastertitel”
- wird immer wieder angereichert
- enthält die Daten, die an Drittsysteme weitergegeben werden

Wie kommt es zu Konflikten?

- beim Einspielen gleicher Titel aus verschiedenen Paketen/Quellen
  - Datenkonflikte, z.B. durch vertauschte Identifikatoren

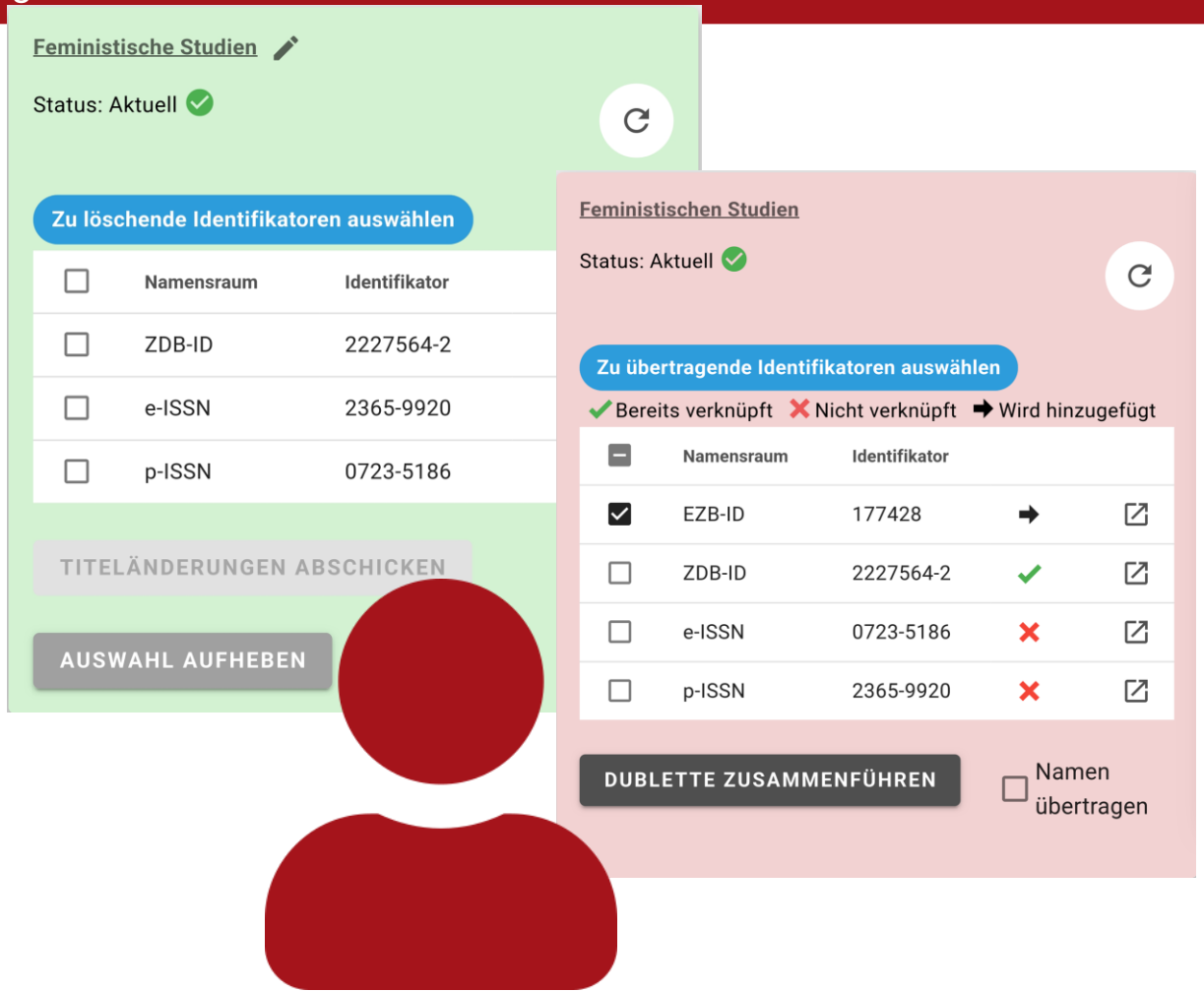


Korrektur


Referenztitel
Feministische Studien
e-ISSN: 2365-9920
p-ISSN: 0723-5186
KONFLIKT!
neues Paket
Pakettitel
Feministische Studien
e-ISSN: 0723-5186
p-ISSN: 2365-9920

Prüfanfragen

- zentrale Workflow-Komponente in der GOKB
- für die manuelle Bearbeitung von Konflikten konzipiert und entwickelt
- Arten von Prüfanfragen:
  - Paketgebunden (z.B. bei Paketupdates)
  - Konflikte mit Referenztiteln
- Kurator*innen als Konfliktlöser
  - community zentriert


Feministische Studien
Status: Aktuell
Zu löschende Identifikatoren auswählen
Namensraum
Identifikator
ZDB-ID
2227564-2
e-ISSN
2365-9920
p-ISSN
0723-5186
TITELÄNDERUNGEN ABSCHICKEN
AUSWAHL AUFHEBEN
Feministischen Studien
Status: Aktuell
Zu übertragende Identifikatoren auswählen
Bereits verknüpft
Nicht verknüpft
Wird hinzugefügt
Namensraum
Identifikator
EZB-ID
177428
ZDB-ID
2227564-2
e-ISSN
0723-5186
p-ISSN
2365-9920
DUBLETTE ZUSAMMENFÜHREN
Namen übertragen

Zentrales Ziel der GOKB: Die Datenqualität an der Quelle verbessern!

Datenqualität verbessern

- Fehler erkennen und korrigieren
  - selbst korrigieren
  - dem Datenlieferanten oder der Kuratorengruppe melden
- Prüfanfragen bearbeiten
  - Daten-Konflikte werden automatisch erkannt
  - zur manuellen Bearbeitung der Daten

Datenqualität verbessern – Wer korrigiert?

- GOKB-Community
  - Kuratoren korrigieren ihre eigene Daten
  - EZB-GOKB Taskforce
    - 30 Kolleg*innen aus verschiedensten Bibliotheken korrigieren in enger Zusammenarbeit mit EZB, ZDB und GOKB Prüfanfragen an den 355 EZB-Paketen
    - ca. 3700 Prüfanfragen wurden 2026 bearbeitet
  - Kolleg*innen aus der ZDB
  - Unterstützung durch das GOKB-Team ist immer gegeben

Melden

Umgang mit gefunden Konflikten I

- Fehler in den KBART-Dateien der Anbieter
  - direkt beim Anbieter melden
  - GOKB stellt Funktionen dafür zur Verfügung
    - Prüfanfragen im Paket exportieren und für Weitergabe an Anbieter vorbereiten
    - Berichte über invalide KBART-Dateien können beim KBART-Validator exportiert werden
- “Fehler in der GOKB”
  - Ursache ausmachen und Datenquelle korrigieren
  - ggfl. in GOKB korrigieren oder GOKB-Team benachrichtigen

Umgang mit gefunden Konflikten II

- Fehler in den Daten der EZB
  - selbst korrigieren (wenn die Möglichkeit besteht)
  - Fehler direkt über die GOKB melden
    - manuelle Prüfanfrage für Kommunikation mit EZB
    - eskalieren von Journal-Prüfanfragen an das GOKB-Team
  - Fehler an die entsprechende Quelle melden

Umgang mit gefunden Konflikten III

- Fehler in der ZDB
  - wenn möglich: selbst beheben
  - Melden!
    - direkt an ZDB
    - an support@gokb.org
    - manuelle Prüfanfrage für Kommunikation mit EZB

Datenbereinigung in der ZDB

Datenbereinigung in der ZDB

1. Korrektur von Identifikatoren
   - e-ISSN
   - p-ISSN
2. Entitäten (Anzahl Titelaufnahmen in der ZDB)
   - Vorhandene Titel nachträglich splitten und Neuaufnahmen für Vorgänger-Titel anlegen
   - Umlenken bei Vorgänger-/Fortsetzungs-Titeln, wenn im ISSN-Portal nur eine ISSN vergeben wurde
     - Änderung der Haupteintragung (sehr selten)
     - Änderung der Erscheinungsjahre

1.1 Korrekturen von Identifikatoren: e-ISSN

- e-ISSN
  - Anstelle einer eigenen e-ISSN ist im Feld 2010 die ISSN der Printausgabe eingetragen

    - 0500 Obvz
      2010 1234-5679*
      4000 Das @Beispiel
      4243 Erscheint auch als$nDruck-Ausgabe!IDN!Das Beispiel. - ISSN 1234-5679

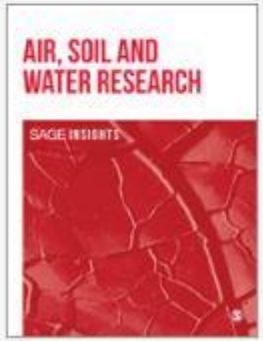
  - Für einen Fortsetzungstitel liegt noch keine eigene e-ISSN vor bzw. in der Ressource ist die e-ISSN des Vorgängers genannt

    - 0500 Obvz
      2010 9876-1234*
      4000 Mein neuer Titel
      4244 f#Fortsetzung von!IDN!Mein Beispiel-Titel. - ISSN 9876-1234

1.1 Korrekturen von Identifikatoren: e-ISSN weiteres Beispiel

Air, Soil and Water Research

Impact Factor: **2.8** / 5-Year Impact Factor: **3.3**



Air, Soil and Water Research is an international, peer reviewed, open access journal focusing on research pertaining to soil, air and water. The journal seeks to report on properties, includin
and also analysis, microbiology,
plants and crops, … | View full jo
This journal is a member of the

Air, Soil and Water Research

ISSN: 1178-6221

Online ISSN: 1178-6221

1.2 Korrekturen von Identifikatoren: p-ISSN

- p-ISSN
  - Aufgrund früher geltender Splitregeln kann es vorkommen, dass Druck-Ausgaben mehr als ein Feld 2010 enthalten

    - 0500 Abvz
      2010 2470-0045*$t2016- → **aktuelle** ISSN steht an erster Stelle
      2010 1539-3755*$t2001-2015 → frühere ISSN
      2010 1063-651X*$t1993-2000 → frühere ISSN
      3220 Physical review / E
      4000 Physical review
      4005 *E*Statistical, nonlinear and soft matter physics
      4025 47.1993 -
      4243 Erscheint auch als$n**Online-Ausgabe, 2016-**!IDN!
      Physical review. E. ISSN: 2470-0053
      4243 Erscheint auch als$n**Online-Ausgabe, 2001-2015**!IDN!
      [Physical review / E]. ISSN: 1550-2376
      4243 Erscheint auch als$n**Online-Ausgabe, 1993-2000**!IDN!
      Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics.
      ISSN: 1095-3787

2. Datenbereinigung in der ZDB - Entitäten

1. Vorhandene Titel nachträglich splitten und Neuaufnahmen für Vorgänger-Titel
2. Umlenken bei Vorgänger-/Fortsetzungs-Titeln, wenn im ISSN-Portal nur eine ISSN vergeben wurde
    - Änderung der Haupteintragung (sehr selten)
    - Änderung der Erscheinungsjahre

2.1. Vorhandenen Titel nachträglich splitten : eine ZDB-ID ist mehrfach in der GOKB vorhanden

2095938-2	Strength and conditioning journal (Current) - Journal Strength and conditioning journal (Current) - Journal Strength and conditioning journal (Current) - Journal

Wie kommt es zu den 1:3 - Einträgen (ZDB - GOKB)?

2010 1533-4295*

2110 2095938-2

4000 Strength and conditioning journal / NCSA, National Strength & Conditioning Association

4025 1.1979 -

4243 Erscheint auch als**$n**Druck-Ausgabe, bis 3.1981/82,2!015886212!--Abxz--National Strength Coaches Association [Tb1]**$B**Verfasser: National Strength Coaches Association journal. ISSN: 0199-610X

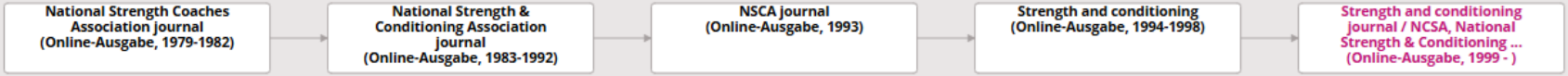
4243 Erscheint auch als**$n**Druck-Ausgabe, 3.1981/82,3 - 15.1993!015886387!--Abvz--National Strength & Conditioning Association [Tb1]**$B**Verfasser: National Strength & Conditioning Association journal. ISSN: 0744-0049

4243 Erscheint auch als**$n**Druck-Ausgabe, 15.1993!01725440X!--Abvz--National Strength & Conditioning Association [Tb1]**$B**Verfasser: NSCA journal. ISSN: 0744-0049

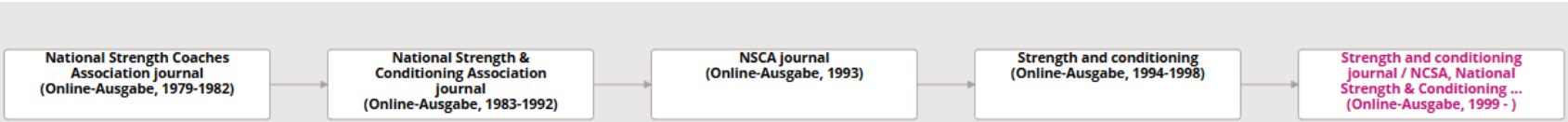
4243 Erscheint auch als**$n**Druck-Ausgabe, 16.1999 - 20.1998!01724532X!--Abvz--: Strength and conditioning. ISSN: 1073-6840

4243 Erscheint auch als**$n**Druck-Ausgabe, ab 21.1999!020399960!--Abvz--: Strength and conditioning journal. ISSN: 1073-6840

Was ist die Lösung: Neuaufnahmen analog ISSN-Portal



Was ist die Lösung? Neuaufnahmen analog ISSN-Portal



3226878-6 im Zeitraum: Volume 1, number 4 (June/July 1979)-volume 4, issue 6 (December 1982)
National Strength Coaches Association journal

3226880-4 im Zeitraum: Volume 5, issue 1 (February 1983)-volume 14, issue 6 (December 1992)
National Strength & Conditioning Association journal

3226882-8 im Zeitraum: Volume 15, number 1 (February 1993)-volume 15, number 6 (December 1993)
NSCA journal

3226883-X im Zeitraum: Volume 16, issue 1 (February 1994)-volume 20, issue 6 (December 1998)
Strength and conditioning

2095938-2 im Zeitraum: 21.1999 -
Strength and conditioning journal

2.2 Umlenken bei Vorgänger-/Fortsetzungs-Titeln, wenn im ISSN-Portal nur eine ISSN vergeben wurde

2.2 Umlenken bei Vorgänger-/Fortsetzungs-Titeln, wenn im ISSN-Portal nur eine ISSN vergeben wurde

ASEE Prism

Identifikatoren

Namensraum	Wert
EZB-ID	237800
ZDB-ID	2549437-5
e-ISSN	1930-6148
p-ISSN	1056-8077

Prism

Identifikatoren

Namensraum	Wert
EZB-ID	237800
ZDB-ID	2549442-9
p-ISSN	1056-8077

Vorgänger-und Fortsetzungstitel mit gleicher p-ISSN

ASEE prism

2549437-5

2010 1930-6148*

4000 ASEE prism / American Society for Engineering Education

4025 1991,Sept. - 1999,4(Apr.); 8.1999,8(Mai/Juni); 9.1999,1(Sept.) - 13.2003/04

4243 Erscheint auch als$nDruck-Ausgabe !1088119-0!--Abvz–:ASEE prism. ISSN: 1056-8077

4244 f#Fortsetzung von!1136638768!--Obvz--: Engineering education news

4244 s#!1001621700!--Obvz--: Prism

Prism

2549442-9

4000 Prism / American Society for Engineering Education

4025 14.2004-vol. 32, no. 3 (2023) ; damit Erscheinen eingestellt

4212 Abweichender Titel bei JSTOR: ASEE prism

4243 Erscheint auch als$nDruck-Ausgabe!026798808!--Abvz--: Prism. ISSN: 1056-8077

4244 f#!1001621379!--Obvz--American Society for Engineering Education [Tb1]$BVerfasser: ASEE prism. ISSN: 1930-6148

ASEE Prism

SEARCH THE JOURNAL

This title is part of a longer publication history. The full run of this journal will be searched.

TITLE HISTORY

1991-2023 • **ASEE Prism**

1974-1991 • Engineering Education News

Prism is the award-winning flagship publication of the American Society for Engineering Education, a nonprofit organization of individuals and institutions committed to furthering education in engineering and engineering technology. The magazine is read by key decision makers within engineering education and reaches engineering faculty members, administrators, and students in all disciplines of engineering.

ASEE Prism

SEARCH THE JOURNAL

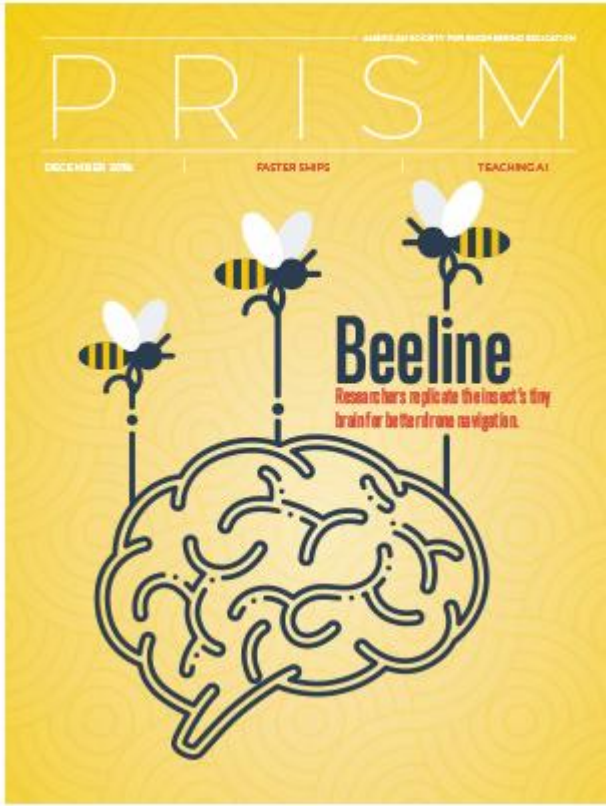
This title is part of a longer publication history. The full run of this journ

TITLE HISTORY

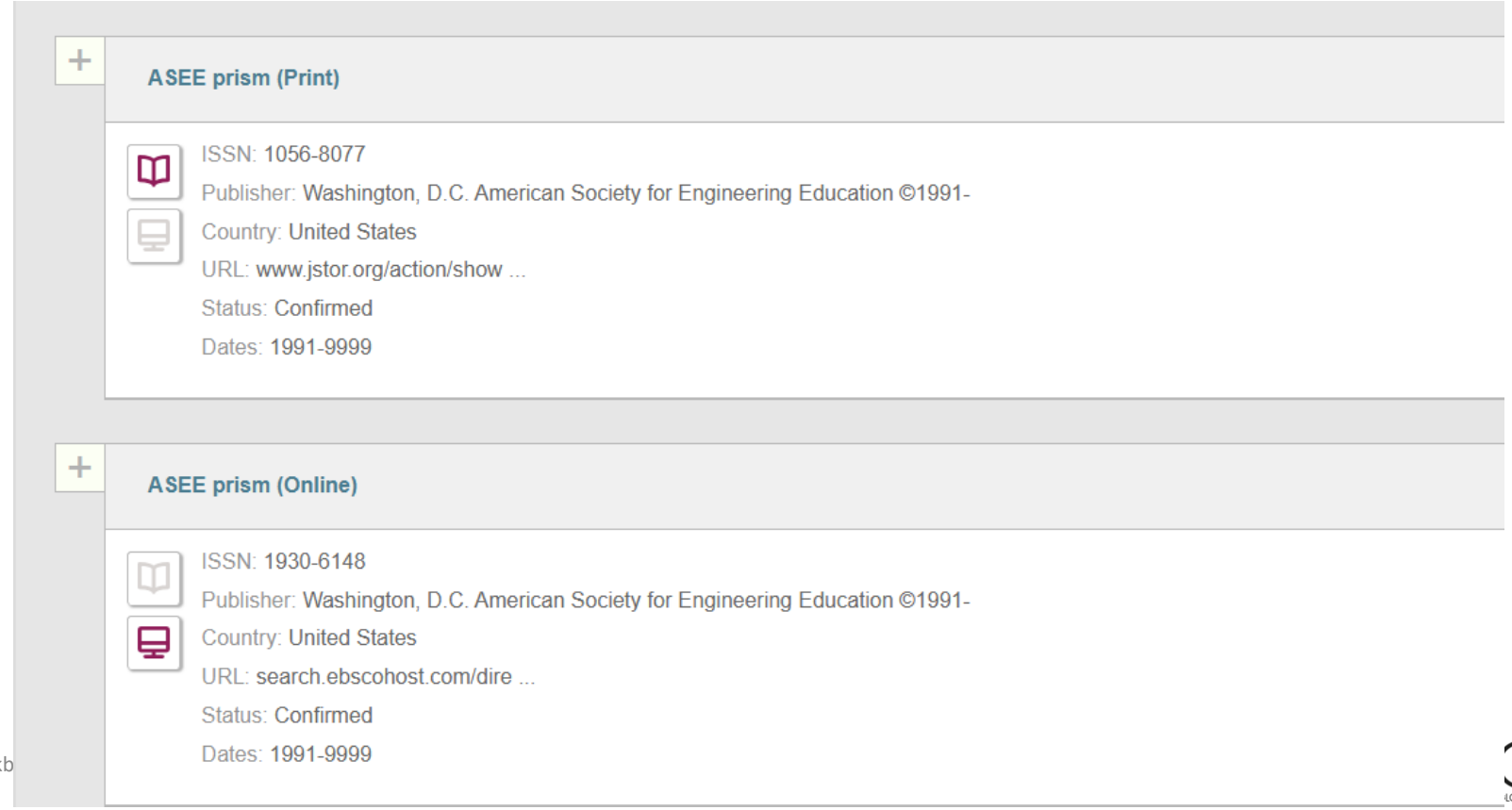
1991-2023 • **ASEE Prism**

1974-1991 • Engineering Education News

Prism is the award-winning flagship publication of the Americ nonprofit organization of individuals and institutions committe engineering technology. The magazine is read by key decision reaches engineering faculty members, administrators, and stu



Blick ins ISSN-Portal



ASEE prism (Print)

ISSN: 1056-8077
Publisher: Washington, D.C. American Society for Engineering Education ©1991-
Country: United States
URL: www.jstor.org/action/show ...
Status: Confirmed
Dates: 1991-9999

ASEE prism (Online)

ISSN: 1930-6148
Publisher: Washington, D.C. American Society for Engineering Education ©1991-
Country: United States
URL: search.ebscohost.com/dire ...
Status: Confirmed
Dates: 1991-9999

2.2 Umlenken bei Vorgänger-/Fortsetzungs-Titeln, wenn im ISSN-Portal nur eine ISSN vergeben wurde

- EZB-Titelhistorie löschen
- Umlenkung in ZDB eintragen
- GOKB-Referenztitel korrigieren

ASEE prism

2549437-5

2010 1930-6148*

4000 ASEE prism / American Society for Engineering Education

4025 1991,Sept. - 1999,4(Apr.); 8.1999,8(Mai/Juni); 9.1999,1(Sept.) - 13.2003/04

4243 Erscheint auch als$nDruck-Ausgabe !1088119-0!--Abvz–:ASEE prism. ISSN: 1056-8077

4244 f#Fortsetzung von!1136638768!--Obvz--: Engineering education news

4244 s#!1001621700!--Obvz--: Prism

Umlenkung

Prism

2549442-9

2010 1930-6148*

4000 Prism / American Society for Engineering Education

4025 Vol. 1, no. 1 (September 1991)-vol. 32, no. 3 (2023) ; damit Erscheinen eingestellt

4212 Abweichender Titel: ASEE Prism

4243 Erscheint auch als$nDruck-Ausgabe!026798808!--Abvz--: Prism. ISSN: 1056-8077

4244 Fortsetzung von!1136638768!--Obvz--: Engineering education news. ISSN: 2573-0592

Datenbereinigung in der ZDB

- Es geht nicht um die Ästhetik von Titelaufnahmen

- Ziel sind
  - **eindeutige Identifikatoren**
  - **international einheitliche Entitäten**

Fazit

Jeder definiert gute Daten anders – aber alle sind sich einig, dass fehlerhafte Daten zu Problemen führen!

Das Ausmerzen von Fehlern an der Quelle verhindert unentdeckte Konflikte in den Bibliotheksmanagement- und Discovery-Systemen, die zu Matching-Problemen führen können.

Die GOKB dient als Werkzeug, um Fehler in den Quelldaten ausfindig zu machen und durch direkte Behebung im System als Schutzschild gegen die Weitergabe von fehlerhaften Informationen.

CC BY

Wir nehmen Ihre Fragen und Anregungen gerne entgegen!

www.gokb.org/