

Datenvergleich

15.06.2021

Anzahl Publikationen

Der Datensatz der Medizinischen Bibliothek enthält **15178 Publikationen** für den Zeitraum 2018 bis 2020.

Der BIH-Datensatz enthält **13252 Publikationen** für den Zeitraum 2018 bis 2020.

Publikationen in den Datensätzen nach Jahren

Auf die einzelnen Jahre des Untersuchungszeitraums bezogen ergibt sich folgende Verteilung:

year	medbib	bih
2018	4421	3906
2019	4676	4329
2020	6081	5017

Überschneidungen und Abweichungen

Vergleich der Datensätze anhand der DOI

Um die Überschneidungen und Abweichungen mit einem eindeutigen Identifier analysieren zu können, wurden nur die Publikationen mit einer DOI verwendet.

Keine DOI hatten 179 Publikationen der Medbib-Daten und 99 Publikationen der BIH-Daten. Diese Publikationen wurden aus den Datensätzen entfernt. Somit wurden 14999 Publikationen aus dem Medbib-Datensatz und 13153 Publikationen aus dem BIH-Datensatz ausgewertet.

Beide Datensätze wurden anhand der DOI miteinander verbunden. Der verbundene Datensatz enthält 15694 Publikationen. 12460 der Publikationen sind in beiden Datensätzen enthalten.

Der Medbib-Datensatz enthält 2541 Publikationen, die nicht im BIH-Datensatz enthalten sind.

Der BIH-Datensatz enthält wiederum 693 Publikationen, die nicht im Medbib-Datensatz enthalten sind.

Die BIH-Daten, die nicht im Medbib-Datensatz enthalten sind, wurden in der Datentabelle in der Spalte `in_bih_dataset` mit logischen Operatoren gekennzeichnet.

Überprüfung des OA-Status

In einem nächsten Schritt wurde der OA-Status überprüft.

981 von den 12460 in beiden Datensätzen enthaltenen Publikationen wurden nicht mit dem identischen Open-Access-Status identifiziert.

Diese Datensätze wurden zur weiteren Überprüfung in der Datentabelle in der Spalte `oa_status_identical_medbib_bih` mit logischen Operatoren gekennzeichnet.

Datenvergleich Medbib und Unpaywall

Der Vergleich der Medbib-Daten mit den Unpaywall-Daten ergab, dass der der Open Access-Status in 282 Fällen nicht identisch ist.

Die Erscheinungsjahre sind in 1905 Fällen nicht identisch.

Die Abweichungen wurden in der Datentabelle in jeweils einer neuen Spalte (`oa_status_identical_medbib_unpaywall` und `year_medbib_unpaywall_identical` mit logischen Operatoren gekennzeichnet.