



Wie wertvoll sind unsere Bibliotheksdaten wirklich?

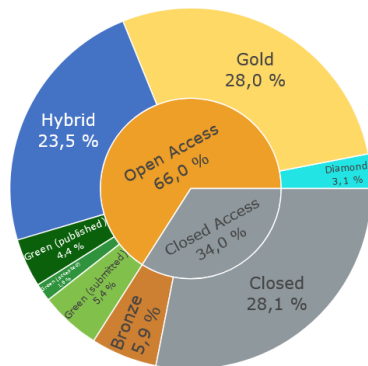
Data is king

Thomas Stäcker (ULB Darmstadt/FH Potsdam)

Berlin 11.7.2025, KOBV-Forum 2025

Open Access – eine Erfolgsgeschichte der Digitalisierung

Verteilung der Zeitschriftenartikel in
Deutschland



Die Grafik zeigt das Open/Closed Access Verhältnis der Zeitschriftenartikel (768.540) in Deutschland für die letzten fünf Jahre (2020 - 2024) auf Basis von OpenAlex, Unpaywall sowie den im OAM verwendeten Zeitschriftenlisten (DOAJ, DOAG) an.

Quelle: OA Monitor

2/3 der
Publikationen
erscheinen OA

Entleihungen physischer
Medien haben sich seit
2017 halbiert

Deutsche Bibliotheksstatistik (DBS) - Variable Auswertung

Trefferanzahl: 5

NR	Anzahl der ber. Bibs	Bibliothekstyp	Nr.:167 Entleihungen						
			2017	2018	2019	2020	2021	2022	2023
1	22	Regionalbibliothek	7.576.292	7.400.803	7.136.167	5.908.267	5.768.034	5.889.889	5.609.089
2	2	Nationalbibliothek	1.545.562	1.433.206	1.346.510	1.012.802	958.439	1.260.495	1.408.584
3	64	Universitätsbibliothek	42.575.406	39.976.775	37.250.590	30.244.203	26.655.320	23.271.459	20.376.968
4	102	Fach-/Hochschulbibliothek	11.965.323	11.328.803	10.717.037	7.774.679	6.529.332	5.879.977	5.478.647
5	2	Zentrale Fachbibliothek	1.558.963	1.427.407	1.282.977	1.041.392	1.011.359	786.708	660.795
Summe			65.221.546	61.566.994	57.733.281	45.981.343	40.922.484	37.088.528	33.534.083
Mittelwert			13.044.309,2	12.313.398,8	11.546.656,2	9.196.268,6	8.184.496,8	7.417.705,6	6.706.816,6

Länderbezug: Deutschland

Fragebogen: Wissenschaftliche Universal- und Hochschulbibliotheken

Die Sammlungssituation in nicht allzu ferner Zukunft

- Weitgehend alle aktuellen wissenschaftlichen Texte stehen online OA zur Verfügung und bedürfen keiner vermittelnden Bibliothek mehr
- Urheberrechtsfreie ältere Werke stehen weitgehend vollständig oder auf Abruf online OA zur Verfügung
- Der Erwerbungsetat wird überwiegend für Publikationen, nicht mehr für Käufe oder Lizenzen aufgewandt
- Archiviert werden Werke aus begrenzten Räumen im Rahmen der gesetzlichen regionalen oder nationalen Pflichtablieferung

Was bleibt für die lokale Sammlung?

- Urheberrechtsbewehrte Werke
- Werke in nicht-digitaler Form
- Pflichtexemplare
- Originale der kulturellen Überlieferung
- ...
 - Reduktion auf Archivfunktion
 - Musealisierung
 - Schwindende wissenschaftliche Relevanz

Sammlung „just in time“

„... die überwiegende Zahl der wissenschaftlichen Bibliotheken [wird]... ihren eigenen Bestand im Sinne einer ganz auf Service abstellenden Gebrauchssammlung, eine ›online core collection‹ für das Tagesgeschäft, ausrichten und letztlich nur noch ›just in time‹ »sammelnd« aktiv werden.“

(Kempf 2022)

Ende des „Gutenberg terror“

“I predict that within ten years, the days of wondering whether or not one’s library can give access to a desired resource will be a fading memory, in much the same way that reliance on a printed card catalog is a faint and fading memory in 2011.”
(Anderson 2011)

Aufstieg der Data Science

LITERATUR RECHNEN

Lektüre im Computerzeitalter

(Lauer 2009)

Always Already Computational: Collections as Data



The
Fourth
Paradigm

(Hey 2009)

(Santa Barbara Statement 2018)



Siegeszug der KI

- https://www.destatis.de/DE/Presse/Pressemitteilungen/2024/11/PD24_444_52911.html
- <https://huggingface.co/>
- <https://osai-index.eu/>

Data Science und KI brauchen Daten, aber ...

- in den letzten zwanzig Jahren haben Bibliotheken digitale Medien nicht gesammelt, sondern nur lizenziert (Nutzungstyp Lesen)
- Kopien von National- und Allianzlizenzen werden berechtigten Einrichtungen nur begrenzt zur Verfügung gestellt
- Archivkopien von lizenzierten digitalen Medien sind für die Wissenschaft weitgehend unzugänglich (NatHosting, u.ä.)



TECHNISCHE
UNIVERSITÄT
DARMSTADT

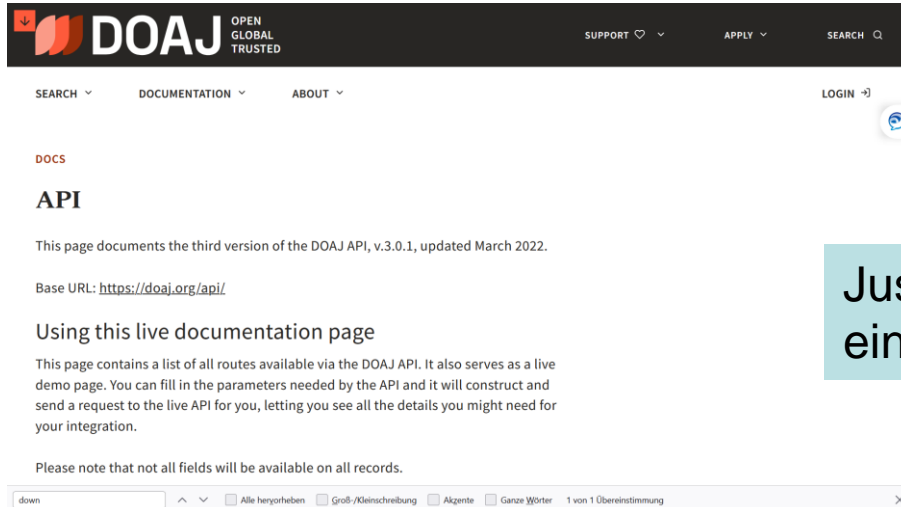
Quelle: <https://pixabay.com/de/vectors/obdachlos-arm-unbehaust-ungesch%C3%BCtzt-295489/>

Sonderfall Katalog

Library Catalogues as Data Research, Practice and Usage Edited By Paul Gooding, Melissa Terras, Sarah Ames Copyright 2025



Ist doch alles im Netz ...



The screenshot shows the DOAJ (Open Access) API documentation page. The header includes the DOAJ logo with the tagline 'OPEN GLOBAL TRUSTED' and navigation links for 'SUPPORT', 'APPLY', and 'SEARCH'. Below the header, there are tabs for 'SEARCH', 'DOCUMENTATION', and 'ABOUT'. The main content area is titled 'DOCS' and 'API'. It states: 'This page documents the third version of the DOAJ API, v.3.0.1, updated March 2022.' The 'Base URL' is given as <https://doaj.org/api/>. A section titled 'Using this live documentation page' explains that it provides a list of all routes available via the API and serves as a live demo page where users can fill in parameters to see the details of a specific record. A note at the bottom says 'Please note that not all fields will be available on all records.' At the very bottom, there is a search bar with the word 'down' and several checkboxes for filtering results: 'Alle hervorheben', 'Groß-/Kleinschreibung', 'Allegente', and 'Ganze Wörter', followed by the text '1 von 1 Übereinstimmung'.

Just in time? Wie komme ich an
ein Dataset der 95338 Artikel?

Das Ende der Sammlung?

- „Just in time“ braucht die Aggregation „just in case“
- Sammeln ist Ordnen innerhalb eines Ordnungsrahmens. Welches ist und wer schafft diese Ordnung?
- Reproduzierbarkeit und Provenienz von algorithmischen Auswertungen
- Datensouveränität



➤ FAIRe Publikationen

Zehn Leitgedanken zum Aufbau digitaler Sammlungen

1. Jede erworbene (lizenzierte oder OA) Publikation wird als Datei erworben. Retrodigitalisierung ist ein Erwerbungsfall.
2. Jede Publikation wird in Originalform gespeichert und mindestens zusätzlich in ein standardisiertes Basisformat konvertiert
3. Erworbene Dateien werden in der Regel dauerhaft nach OAIS Prinzipien archiviert
4. Alle relevanten Metadaten werden aus den Daten extrahiert und (semi-)automatisch standardisiert
5. Standards sind webtaugliche Standards und berücksichtigen die Maschinenlesbarkeit der Publikationen
6. Publikationen werden nach Erwerbungsrichtlinien gesammelt (gefiltert)
7. Publikationen lassen sich nach inhaltlichen, formalen und transformativen Kriterien zu frei konfigurierbaren Datensets zusammenstellen und abrufen.
8. Datensets von OA Publikationen sind auch OA
9. Datensets können zum Zwecke der Reproduzierbarkeit persistent adressiert werden
10. Digitale Sammlungen können als ganze nachgenutzt werden.

Prinzipien der bibliothekarischen Zusammenarbeit

1. Alle können alles haben. Keine Unikalität mehr von Beständen
2. Kooperative Aufbereitung der Publikationen und komplettes Datenset aller wissenschaftlichen Publikationen in einem zentralen Verbund oder Verbünden nach dem Modell der Verbundkataloge.
3. (Lokale) Nutzung der Publikationen nach dem Modell der „Titelkopie“.

„But fundamentally, you can stick a dokument in any of these archives
and it will get replicated to all the other archives“
(Hey et al 2009)

Leistungen der einzelnen Bibliotheken

- Herstellen von FAIRen Publikationen
- Aggregieren und Vereinheitlichen (*data cleaning*) von (externen) Sammlungen
- Selektion von Teilen einer Sammlung
- Transformieren in neue Derivate („KI ready“)
- Extraktion von Wissen / Informationen aus Sammlungen
- Langzeitarchivierung und persistente Zugänglichkeit
- ...

„Wie wertvoll sind unsere Bibliotheksdaten wirklich?“

- **Sehr wertvoll, wenn wir sie denn gesammelt haben**

»Die meisten *Bibliotheken sind **entstanden***, nur wenige sind **angelegt** worden, deren sich ein so kundiger Fürst, als August war, in einer **ununterbrochenen Folge** von nahe fünfzig Jahren beeiferte.“
(Lessing 1773)

Vielen Dank für Ihre
<emph>Aufmerksamkeit!</emph>

