

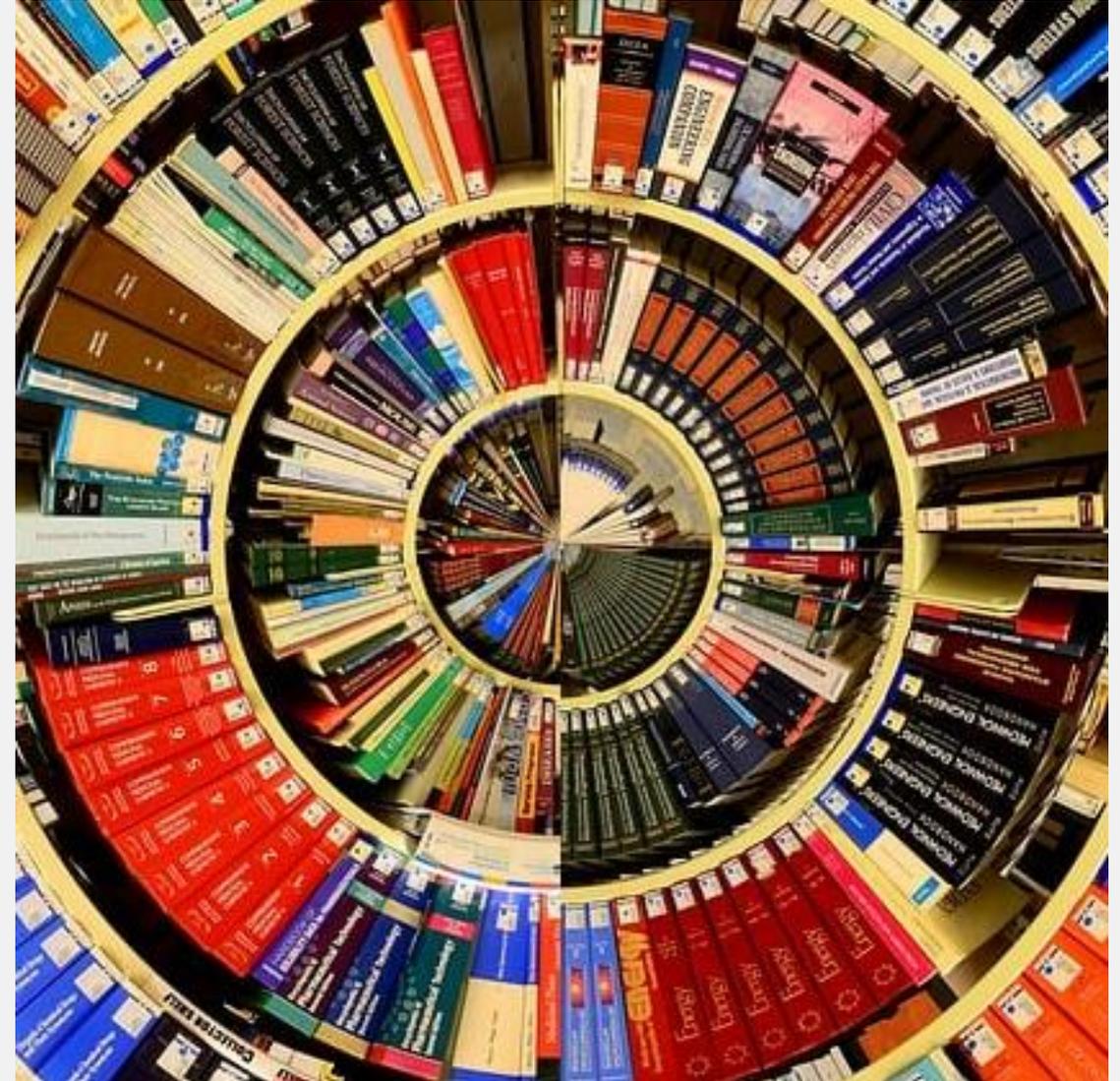
Ergeben Publikationsnachweise in ORCID einen Mehrwert für akademische Suchmaschinen?

111. BiblioCon 2023, 25. Mai 2023

 [Jochen Schirrwagen, Universität Bielefeld - UB](#)

 [Friedrich Summann, Universität Bielefeld - UB](#)

 [Dirk Pieper, Universität Bielefeld - UB](#)



Kontext: Im Rahmen des DFG-Projekts ORCID-DE 2 (2019 - 2022)

Projektpartner:innen



Gefördert durch



Initiiert durch



<https://www.orcid-de.org/ueber-orcid-de/orcid-de-projekt>

Hintergrund & Motivation

- Bestandsaufnahme der bibliographischen Metadaten in der ORCID-Registry
- Prüfung der Metadatenanreicherung in BASE
- Optimierung der BASE-ORCID “Kommunikation” auf Basis bestehender Services, wie
 - BASE Search & Link Wizard
 - ORCID-DE Monitor, <https://monitor.orcid-de.org/>



Aufbau des ORCID-Record Bereichs Activities



Activities Sections

- Distinction
- Education
- Employment
- Invited Position
- Membership
- Qualification
- Service
- **Works**
- Funding
- Peer review
- Research Resource

Vgl. <https://info.orcid.org/documentation/integration-guide/orcid-record/#ACTIVITIES>

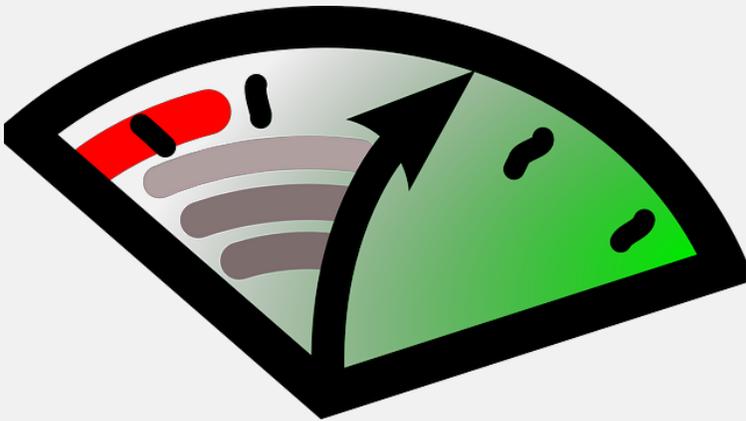
Verwendete Datenquellen und Vorgehen

- ORCID Public Data File 2021. ORCID. Dataset. <https://doi.org/10.23640/07243.16750535.v1>
 - Transformation der XML Works-Daten nach CSV und Analyse
- BASE Dump vom 07/2022 mit 315 Mio Records
 - Überführung in eine key-value Struktur mit DOIs als key und BASE-Nachweise als values
- Unpaywall Dump von 03/2022 zur Abschätzung der Open-Access-Anteile und -Stati

ORCID Public Data File 2021

- Betrachteter Zeitraum der zu ORCID Records hinzugefügten Forschungsaktivitäten: 2013 – 2021 (wobei 2013 und 2021 nicht als volle Jahre zu werten sind)
- Insgesamt 14.847.177 ORCID iDs
- Für **3.185.421** ORCID iDs wurden **72.977.711** Metadaten-Einträge zu Werken festgestellt
- Für **53.021.691** Werke wurde eine DOI angegeben, davon sind **24.358.900** DOIs eindeutig
- Werk-Dubletten in ORCID entstehen durch Autorisierung einer Vielzahl an ORCID-Client-Services u. Search & Link Wizards, sowie Mehrfacherfassungen im Fall von Co-Autorenschaften
- Anzahl Werke je ORCID-iD: min: 1; max: 21.398; Durchschnitt: 22,91; Median: 5

Metadaten- Füllstand in ORCID-WORKS



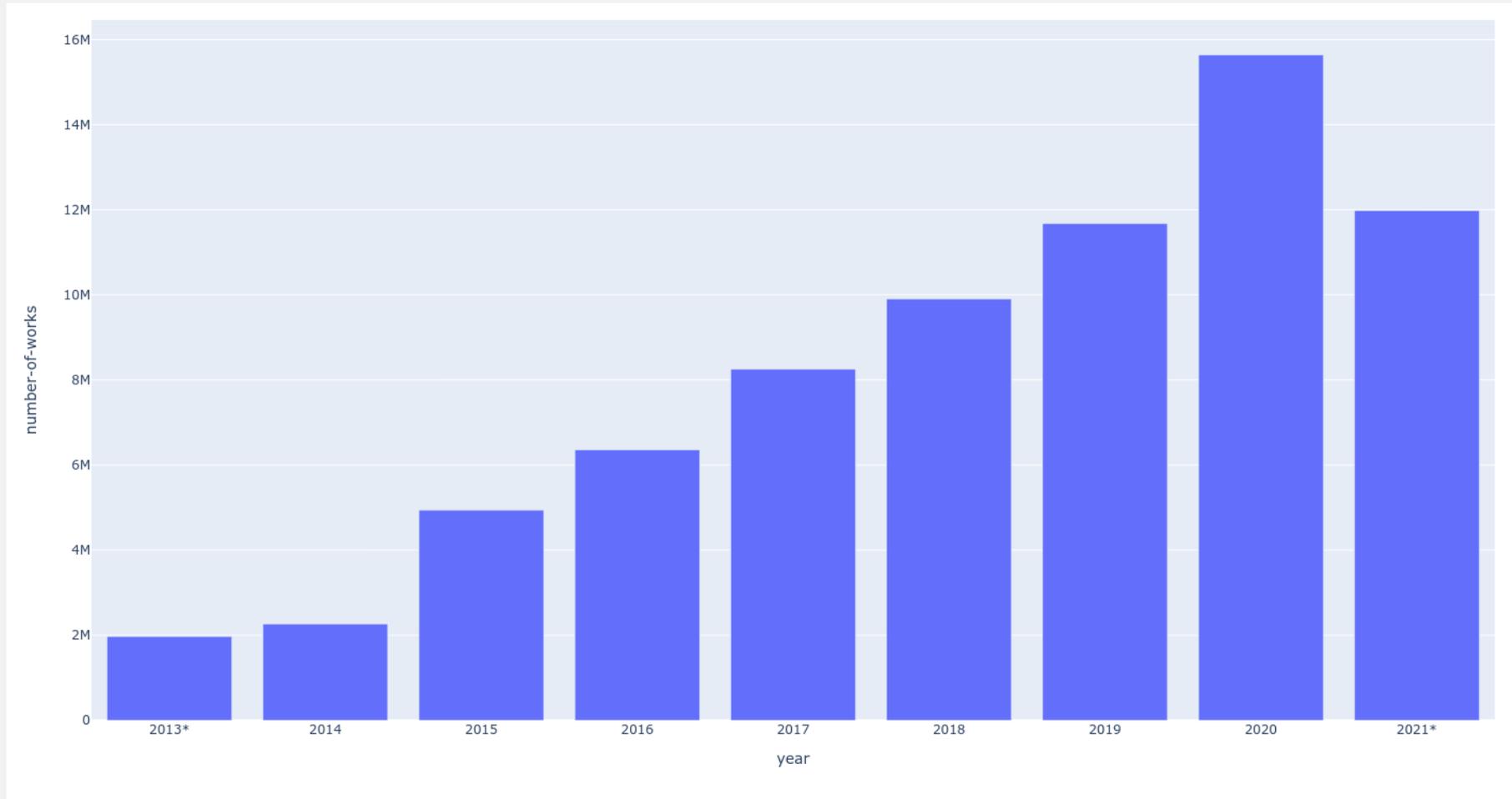
field name	requirement level	no value	fill level (%)
title	mandatory	3,837	99.99
subtitle	optional	71,314,387	2.28
translated-title	optional	72,673,052	0.42
journal-title	optional	8,360,472	88.54
contributor (author)	mandatory (implicit)		100.00
external-id	mandatory for client-apps; optional for ORCID iD user	7,609,682	10.43
external-id-type: doi (of the work)		19,956,020	72.65
(publication) year	optional	1,177,573	98.39
(work) type	mandatory		100.00
url	optional	16,510,379	77.38
short-description	optional	67,446,785	7.58
language-code	optional	n.a.	n.a.
country	optional	n.a.	n.a.
citation	optional	n.a.	n.a.

Erfassung / Import von Metadaten über Werke

- Nur zwei Felder sind verpflichtend:
 - work type und work title
- Erfassungswege:
 - per DOI und Metadaten-Pull aus DOI-Registrierungsagentur
 - per PubMed ID und Metadaten-Pull aus PubMed
 - per Import eines BibTeX-Files
 - Erfassung im ORCID-Formular
- 15.510.251 Einträge durch 1.447.943 ORCID iD Nutzer*innen
- Automatisierte Meldung an ORCID durch ca. 500 Client-Applikationen und Search & Link Wizards

The screenshot shows a web form titled "Works" with a close button (X) in the top right corner. The form is divided into sections. The "Work details" section is highlighted with a green line and includes a red asterisk indicating required information. It contains two red-bordered input fields: "Work type*" (a dropdown menu with "Select a work type" selected) and "Work title*" (a text input field). Below these are "Work subtitle" (text input), "Journal title" (text input), and "Publication date" (Year, Month, and Day dropdowns). A "Link" field is partially visible at the bottom. On the right side, there is a vertical list of blue links: "Work details", "Citation", "Identifiers", "Contributors", "Other information", and "Visibility". At the bottom of the form are two buttons: "Save changes" (blue) and "Cancel" (white).

Zeitlicher Verlauf der an ORCID gemeldeten Werk-Metadaten

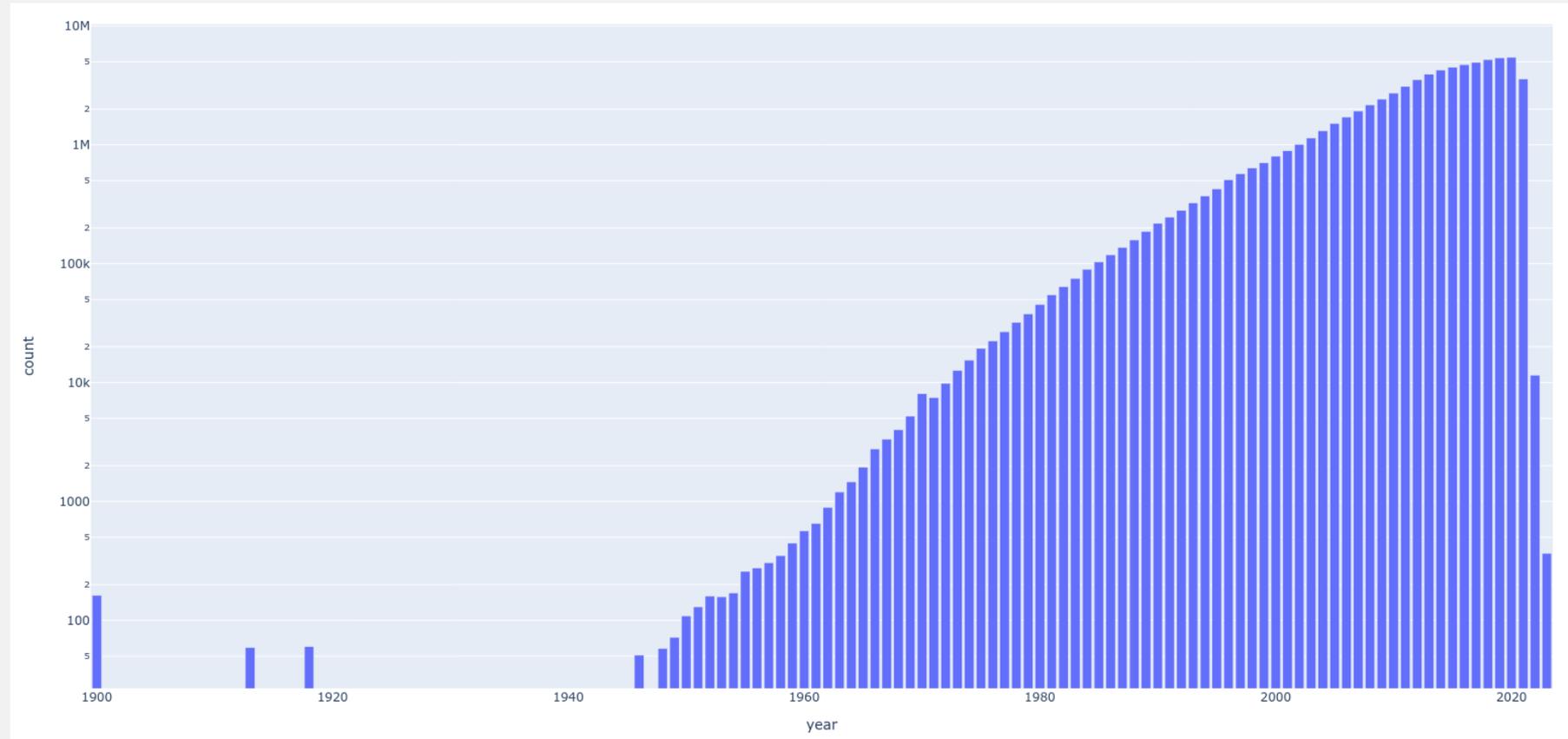


Verteilung über Publikationsjahre

- Keine Festlegung auf bestimmtes Kalendersystem
 - Fehlinterpretation möglich
- Überwiegend junges Publikationsalter

Ausreißer:

Publikationsjahr	Anzahl
Keine Angabe	1.177.572
0	3030
1	49
20	61
1385 – 1440	1590
9999	143

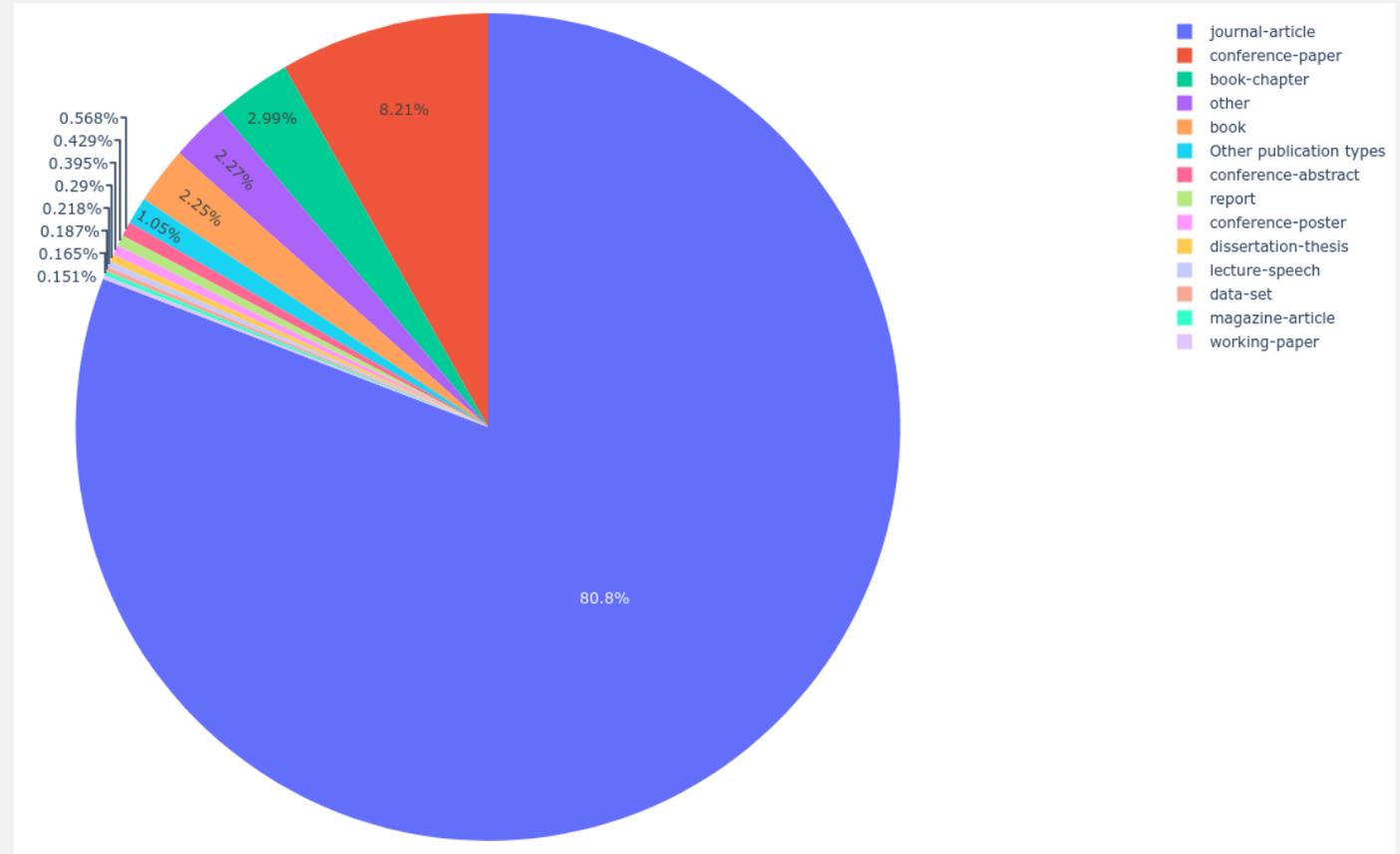


Verteilung der ORCID work types

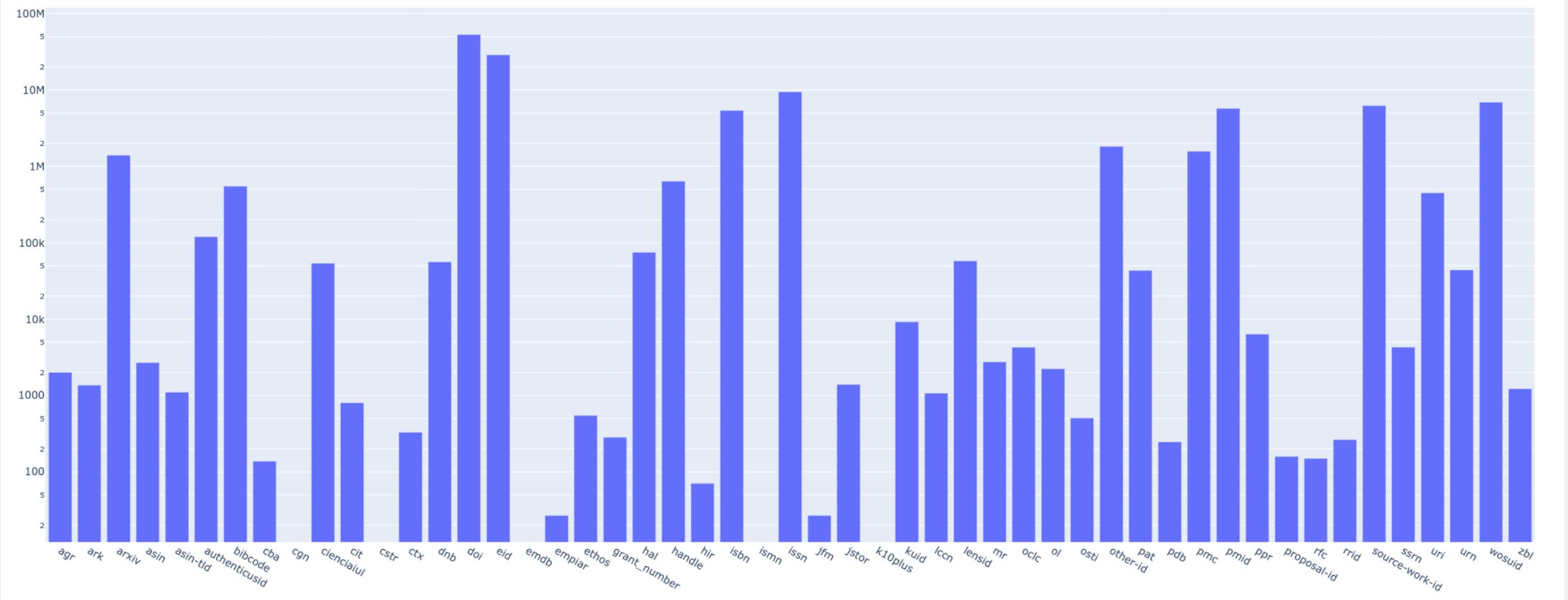
- Ausdifferenziertes type-Vokabular für
 - Publikationen
 - Konferenzbeiträge
 - Geistiges Eigentum
 - Weitere
 - darunter DMPs, Datasets, Software

Vgl. Liste der ORCID work types:

<https://info.orcid.org/ufaqs/what-work-types-does-orcid-support/>



Häufigkeit von Identifier-Typen in ORCID works

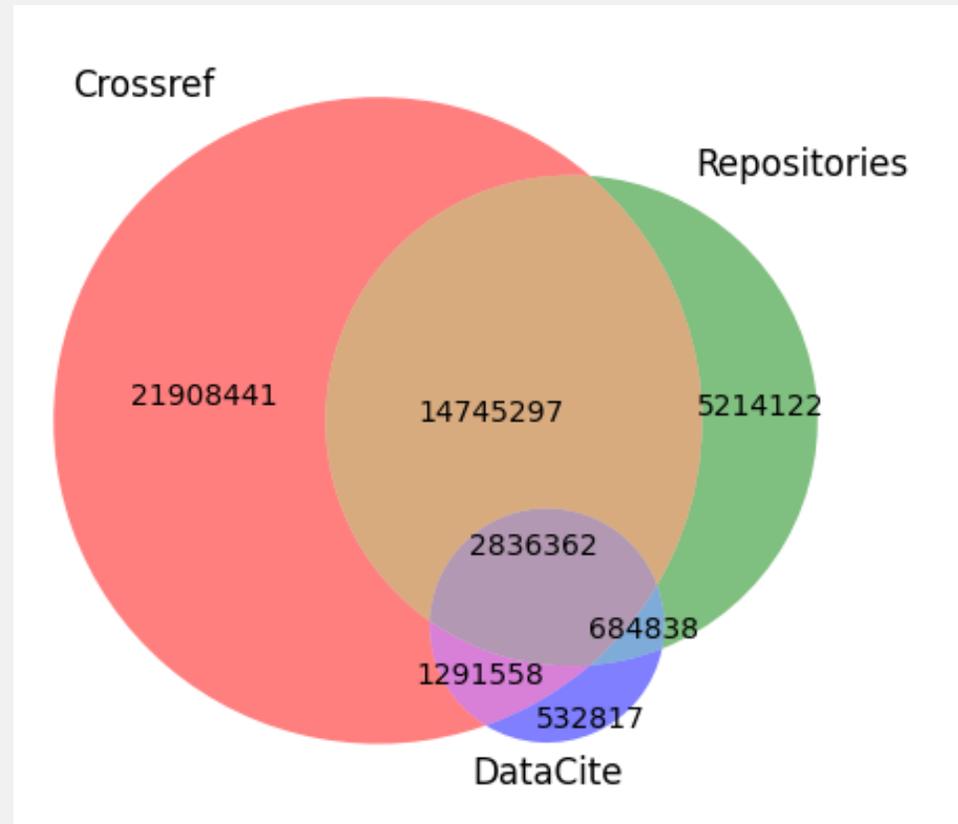


Vgl. Liste der ORCID Identifier Types, <https://pub.orcid.org/v3.0/identifiers>

Abgleich der ORCID Works-Metadaten mit BASE

- Grundlage: ~317 Mio BASE-Metadaten aus über 10.000 Quellen
- Unterteilung der BASE-Metadaten in drei Nachweisräume
 - Repositorien (inkl. Preprint-Server, PubMed, eJournals, digitale Sammlungen, Forschungsdatenplattformen): ~ 186,897 Mio
 - Crossref: ~ 95,704 Mio
 - DataCite mit starken Überschneidungen zu Repositorien: ~ 32,856 Mio
- Abgleich mit Works-Metadaten mit DOI
 - Sätze mit mind.1 Nachweis in BASE: 47.214.171
 - Sätze ohne Nachweis in BASE: 5.807.520
 - Pro DOI können in BASE mehrere Nachweise vorliegen, insg. 188.072.438

Aufschlüsselung der Nachweise über die BASE-Nachweisräume



OA-Anteil der Werke mit DOI auf Basis von unpaywall

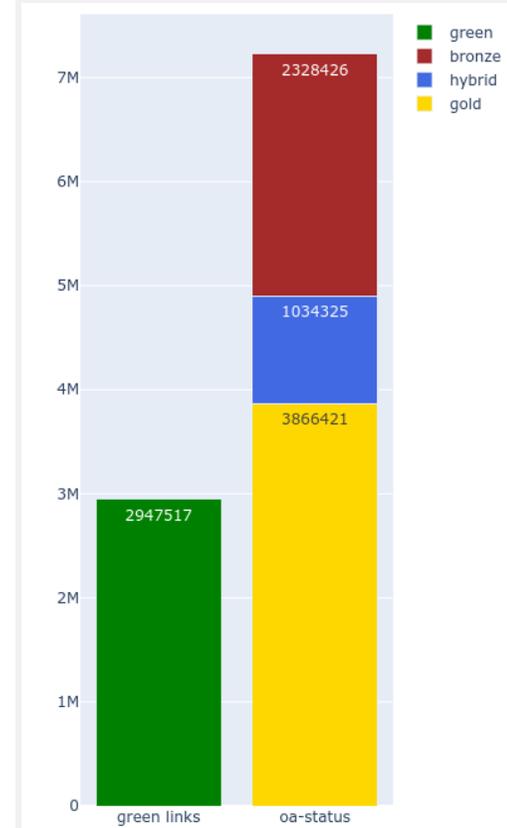
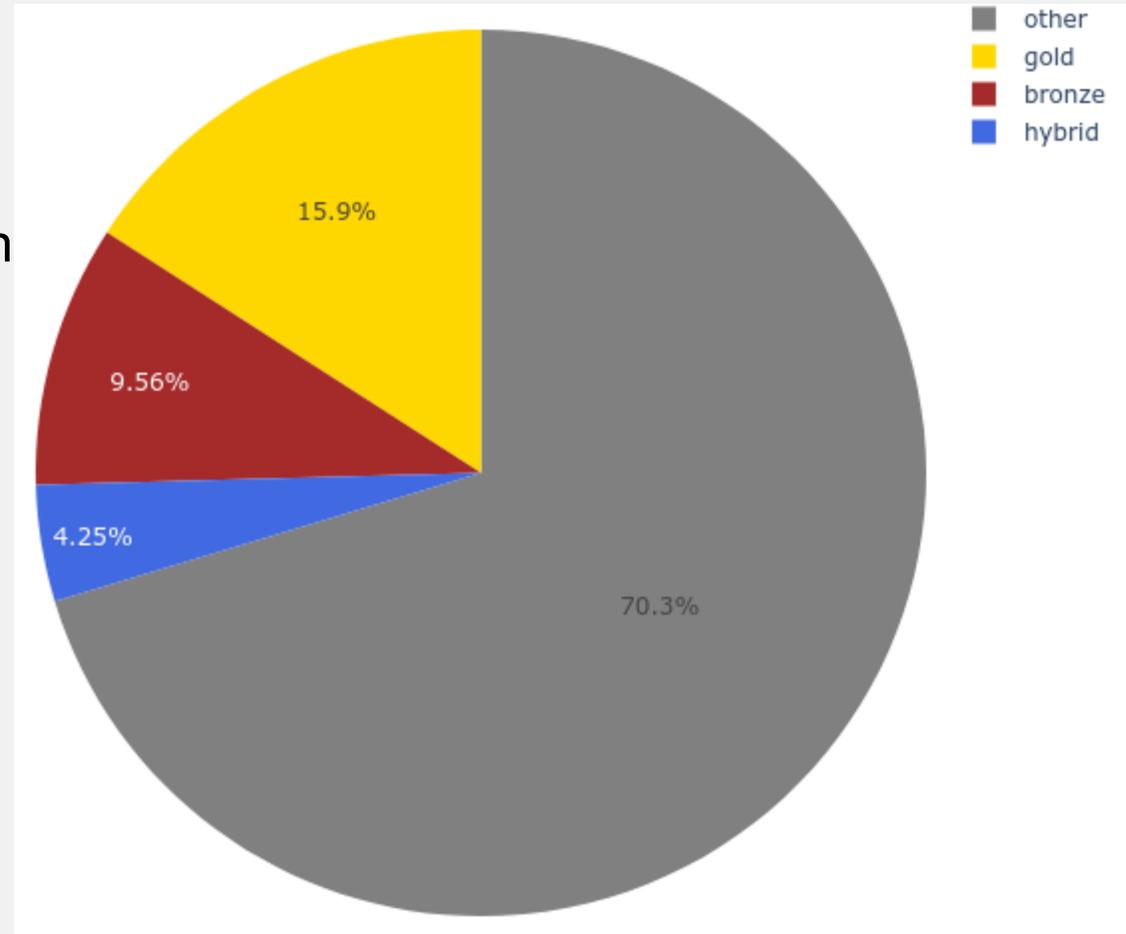
Bestimmung der OA-Stati und
green-Links zu den ORCID Werken

Einschränkung:

- unpaywall berücksichtigt nur
Crossref-DOIs

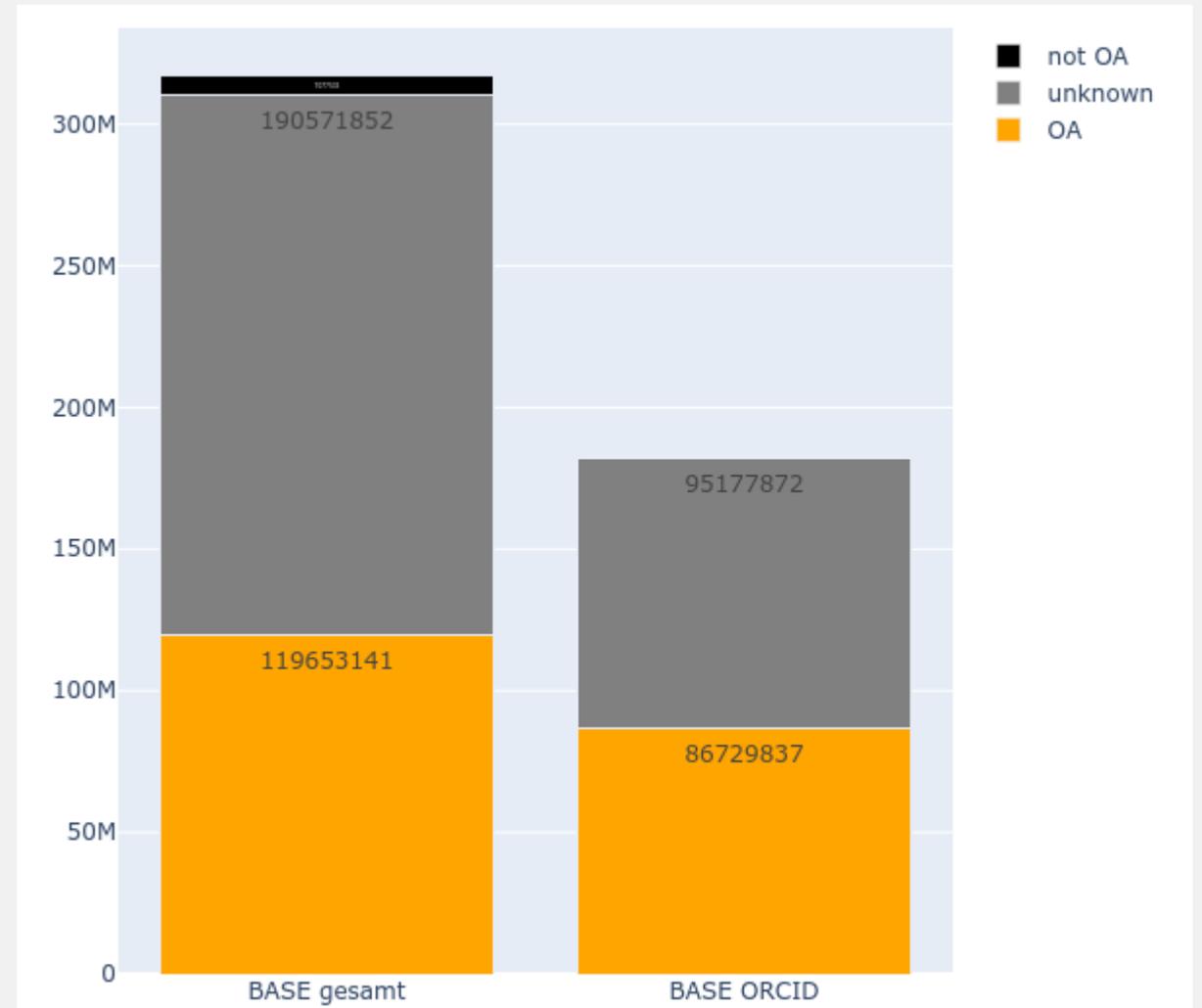
OA-Anteil bei den DOI-

Publikationen mindestens **29,7%**

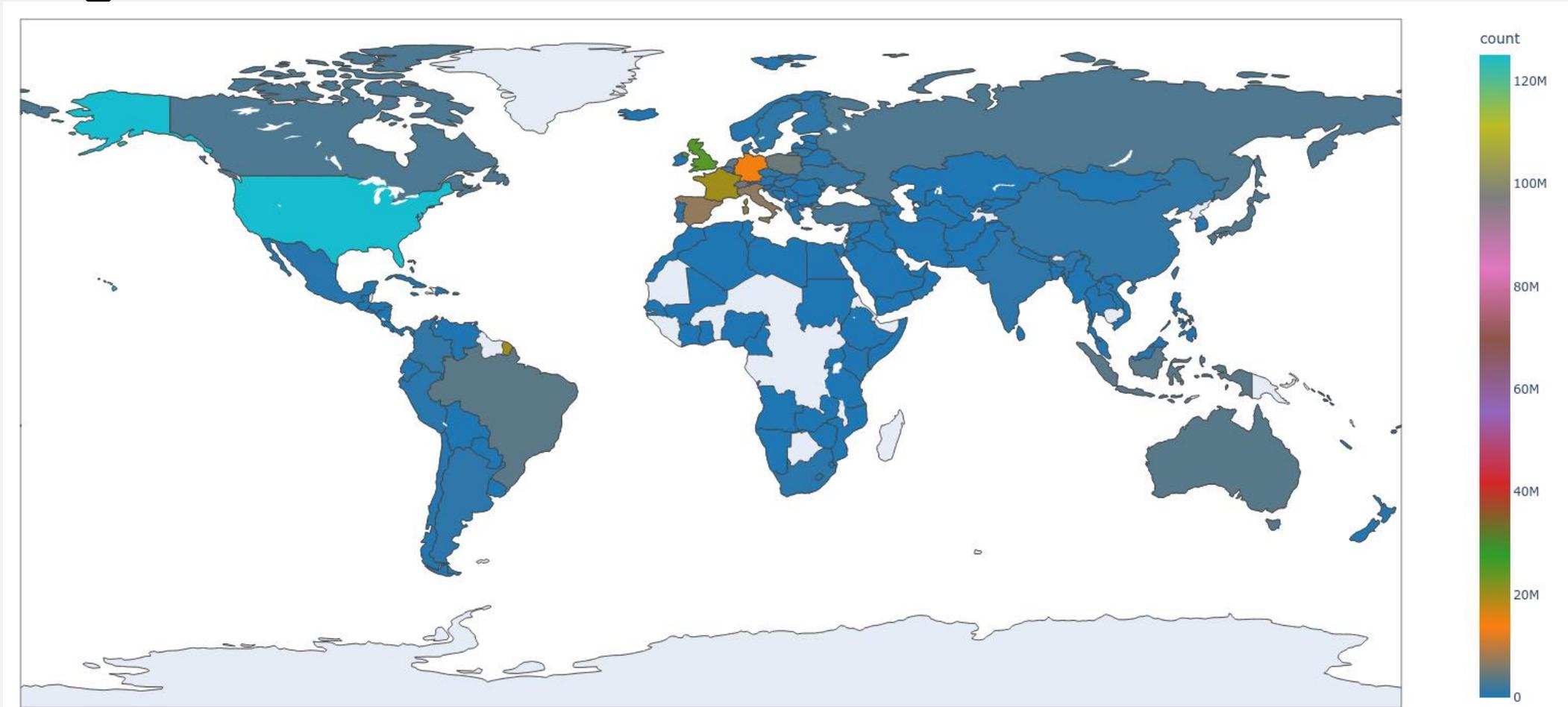


Auswertung der OA-Anteile in BASE

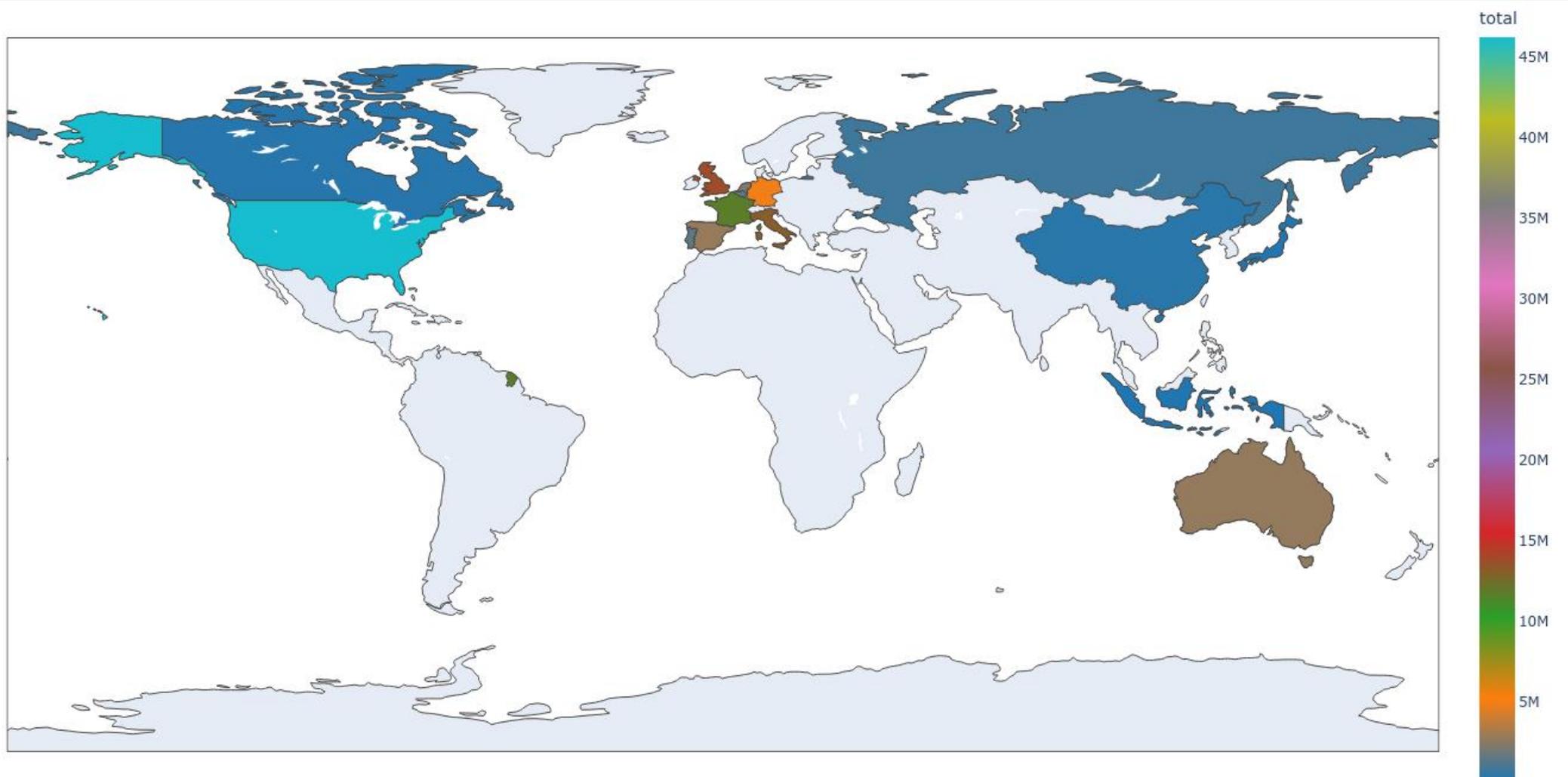
- Aufgrund weiterer Nachweise (Versionen) zu DOI-Publikationen ergibt sich folgendes Bild im Vergleich BASE gesamt und BASE ORCID:
- OA-Anteil BASE gesamt: **37,71%**
- OA-Anteil BASE ORCID: **47,68%**



Geographische Verteilung der Publikationsquellen für BASE gesamt



Geographische Verteilung der Publikationsquellen für BASE-ORCID Nachweise



Anreicherungspotential von ORCID Works-Metadaten in BASE

Gemeinsame Grundlage von Metadatenfeldern in ORCID und BASE	Anreicherungspotential in BASE
Autoreninformationen	ORCID iD
Titel	Zstzl. Titelinformationen (subtitle, translated title)
Publikationsjahr	
Abstractinformationen	Weitere Abstractdaten
Link / URL	
Identifier	Weitere Identifier
	Titel des übergeordneten Werks / Journal-Titel

Erkenntnisse & Schlussfolgerungen im Vergleich ORCID Works und BASE

- Umfang der Metadatenmodelle und kontrollierter Vokabulare ist in den Systemen verschieden
- Sehr große Heterogenität der Erfassungswege und Erfassungsqualität in ORCID
- Nur minimale Pflichtangaben für Werk-Metadaten in ORCID notwendig
- Fehlen von Zugangs- und rechtlichen Informationen in ORCID Werk-Metadaten
- Nachnutzung von Werk-Metadaten ohne Identifizierungsangaben sehr problematisch
- Nachnutzung von Werk-Metadaten mit Identifier erfordert i.d.R. Anreicherung aus weiteren Datenbanken
- Strategien für den Umgang mit sehr großen Dublettenmengen erforderlich

Mehrwerte für Aggregatoren wie z.B. BASE

- Import von Werk-Metadaten aus ORCID in BASE zur Ergänzung bisher nicht vorhandener Nachweise in BASE
- Verbesserung der Autor*innendisambiguierung + ggf. Affiliationsinformationen in BASE Metadaten mit Hilfe der ORCID iDs
- Anreicherung bestehender BASE Metadaten insbesondere bzgl. weiterer Identifier
- Auffinden von Open-Access-Nachweisen zu ORCID Werken
- Unterstützung des BASE Claiming Service (Search & Link Wizard)
- Input für den ORCID-DE Monitor bzw. geplanten PID-Monitor

Vielen Dank für Ihre Aufmerksamkeit.

Fragen und Anmerkungen gerne an:

- Jochen Schirrwagen, jochen.schirrwagen@uni-bielefeld.de
- Friedrich Summann, friedrich.summann@uni-bielefeld.de
- Dirk Pieper, dirk.pieper@uni-bielefeld.de