How open are hybrid journals included in nationwide transformative agreements in Germany?

May 27, 2022

Assessing the volume and share of open access articles in hybrid journals is crucial for the planning and implementation of [transformative agreements](https://esac-initiative.org/about/transformative-agreements/), an evolving type of contracts between libraries and publishers where subscription spending is repurposed for open access publishing (Schimmer, Geschuhn, and Vogler 2015). However, library consortia that mainly negotiate such agreements with large commercial publishers often lack these data. Here, we present [hoaddata](https://subugoe.github.io/hoaddata/), an experimental R package, in which openly available journal-level data about nationwide transformative agreements in Germany are combined with article-level open access status information and country affiliations. Accordingly, hoadata provides essential data for the monitoring and benchmarking of transformative agreements across hybrid journal portfolios and countries.

Interacting with data through R packages like hoaddata makes data analytics more transparent because R packages meet generic principles for computational reproducibility: coherent file organisation, separation of data, methods and results, and specification of the computational environment (Marwick, Boettiger, and Mullen 2018). For data science practitioners, R packages, thus, provide a reliable way to re-use data and code. In our specific case, hoaddata not only contains datasets about hybrid open access. It also comprises code used to compile the data by interfacing our cloud-based [Google Big Query data warehouse](https://subugoe.github.io/scholcomm_analytics/data.html), where we store open scholarly data from [Crossref](https://www.crossref.org/), [OpenAlex](https://openalex.org/) and [Unpaywall](https://unpaywall.org/). hoaddata is automatically built and updated with GitHub Actions. Each merge event into the main branch triggers the execution of code, ensuring that the most recent snapshots available in our data warehouse will be used to obtain data about nationwide transformative agreements. Data changes including updates will be incorporated in the package and tracked with Git, which makes it possible to reproduce different version of the data contained in hoaddata.

In this blog post, we describe the data analytics workflow behind hoaddata. The main purpose of hoaddata is to ship data for monitoring dashboards about the progress of nationwide transformative agreements, which we currently develop in the [HOAD project with the support of the Deutsche Forschungsgemeinschaft](https://gepris.dfg.de/gepris/projekt/416115939), but everyone can install the package from GitHub and use it in R. To demonstrate its potential, we will use hoaddata to illustrate the current state of the transition of hybrid journals to fully open access relative to those journals, which are included in nationwide transformative agreements in Germany.

## Data and methods

hoaddata focuses on nationwide transformative agreements in Germany. As a first step, we draw on the work of the [German Open Access Monitor (OAM)](https://open-access-monitor.de/) to obtain a list of journals under these agreements (Pollack et al. 2022). We merged all journals into a single data file and enriched it with missing ISSN variants. Because publishers register journal-level metadata in Crossref when they first deposit metadata for a given journal including ISSNs, we furthermore matched the OAM journal list with [Crossref’s title list](https://www.crossref.org/documentation/reports/browsable-title-list/) to link ISSN variants to journals as they were represented in Crossref metadata.

This [thread](https://community.crossref.org/t/parallel-titles-for-a-given-issn/2183) is helpful to better understand journal curation workflows in Crossref.

After obtaining a list of hybrid journals linked to nationwide transformative agreements, we determined the article volume by journal and year using Crossref. Following Unpaywall’s approach, Crossref metadata records considered as front matter were excluded.[[1]](#footnote-30) Open access articles were identified through Creative Commons license URLs in Crossref metadata records. License URL were mapped to the different license versions like CC BY.

Because country affiliations are a key data point for nationwide transformative agreements, we used OpenAlex to determine the country share per journal and publisher portfolio. To our knowledge, OpenAlex does not provide information about corresponding authors and their affiliation. Instead, we made use of first-author affiliations. A first author is often regarded as being the [lead author](https://en.wikipedia.org/wiki/Lead_author) who has usually undertaken most of the research presented in the article, although author roles can vary across disciplines. In case OpenAlex did not record any country affiliation, we extracted country names from the metadata field display\_name using regular expressions. We applied full counting to account for multiple country affiliations.

As a result, hoaddata provides the following datasets:

* [oam\_hybrid\_jns](https://subugoe.github.io/hoaddata/reference/oam_hybrid_jns.html): Hybrid journals included in the Open Access Monitor. Data were gathered from Pollack et al. (2022), validated and mapped to Crossref-indexed journals.
* [cc\_jn\_ind](https://subugoe.github.io/hoaddata/reference/cc_jn_ind.html): Prevalence of Creative Commons license variants by year and hybrid journal as obtained from Crossref.
* [cc\_openalex\_inst\_jn\_ind](https://subugoe.github.io/hoaddata/reference/cc_openalex_inst_jn_ind.html): First author country affiliations per journal, year and Creative Commons license. Country affiliations were gathered from OpenAlex.

Article-level data

* [cc\_openalex\_inst](https://subugoe.github.io/hoaddata/reference/cc_openalex_inst.html): Article-level affiliation data from first authors as obtained from OpenAlex. Covers only open access articles under a Creative Commons license in a hybrid journal.

hoaddata can be installed from [GitHub](https://github.com/subugoe/hoaddata):

# install.packages("remotes")  
remotes::install\_github("subugoe/hoaddash", dependencies = "Imports")

You can also directly download the data as csv files from [GitHub](https://github.com/subugoe/hoaddata/). The files stored in the [data-raw](https://github.com/subugoe/hoaddata/tree/main/data-raw) folder of the package together with the code used to create the datasets. Specific SQL queries can be found in [inst/sql/](https://github.com/subugoe/hoaddata/tree/main/inst/sql).

## Use-Case: Open access uptake in hybrid journals included in nationwide transformative agreements in Germany

At the time of writing, hoaddata comprised information about 5,562 hybrid journals included in twenty consortial transformative agreements in Germany. Since 2017, these journals have published 348,978 open access articles with an Creative Commons License, representing an share of 12%.

### Open access uptake in Germany’s nationwide transformative agreements in 2021

Using OpenAlex’s affiliation information, we can break down the performance of transformative agreements to countries. Showing the publication year 2021, the following interactive table compares the global publication volume including open access with that of German lead authors by hybrid journal portfolio. It highlights the dominant position of transformative agreements negotiated by the [DEAL consortium](https://deal-operations.de/) in terms of articles published and open access. It is important to note that [DEAL has reached no agreement with Elsevier](https://www.projekt-deal.de/elsevier-news/) so far, one of the largest scholarly publishers, which is, in turn, not included in the analysis.

Interestingly, Germany’s open access share in 2021 performed in most cases below 80%, suggesting that not all authors made use of open access options, or were not eligible to publish open access, likely because their institution was not part of a consortium. But also some article types might no be included in agreements. For instance, many article types in medical specialist journals from Springer Medizin, which are targeted at medical practitioners, are not covered by DEAL.[[2]](#footnote-44) We were not able to obtain Creative Commons license information from Crossref metadata records representing hybrid journal portfolios from ACM, AIP, Hogrefe and SPIE.

The table furthermore shows large discrepancies between the global open access uptake in hybrid journals and Germany’s. In 2021, 16% of articles published in hybrid journals included in nationwide transformative agreements in Germany were open access, while the open access percentage among articles published by lead authors based in Germany was 65%. Overall, 20% of open access articles published in hybrid journals included in transformative agreements in 2021 were from Germany, although its total publication volume accounted for 4.9% of articles published.

### Country overview

As can be seen from the following figure, which highlights the Top 20 most productive countries, open access uptake in hybrid journals included in nationwide transformative agreements in Germany varies across countries. Notably, lead authors based in the US, China or India – which together accounted for roughly 40% of articles published – did utilize open access options to a much lesser extent than authors from European countries. Together with Germany, the UK, the Netherlands, Sweden and Switzerland have gained an open access share above 50% over the years, most likely because of providing nationwide transformative agreements with similar journal coverage.

Percentage of articles with a Creative Commons license per country. Showing the Top 20 most productive countries in terms of articles published by lead authors between 2017 and 2022 in hybrid journals included in nationwide transformative agreements in Germany. Country-specific charts are sorted by the total number of lead author articles. Data sources: Open Access Monitor Zeitschriftenlisten (v2), Crossref, OpenAlex.

## Discussion and outlook

Over the past years, Germany’s library consortia sucessfully negotiated transformative agreements with commercial publishers, resulting in an increase of open access articles from lead authors based in Germany that were published under an Creative Commons license. But this growth of open access is neither balanced across journal portfolios nor across countries.

As illustrated, there are substantial variations across journal portfolios in terms of open access articles published by lead authors based in Germany. They can be explained by different agreement terms such as the number and types of institutions involved or the restriction of open access publishing options to specific article types. In its recent [Reference Guide to Transformative Agreements](https://esac-initiative.org/about/transformative-agreements/reference-guide/implementation/), and in line with previous research (Borrego, Anglada, and Abadal 2020), the [ESAC initiative](https://esac-initiative.org/) points out that agreement terms affecting the scope of contracts can have a large impact on the performance of transformative agreements.

In terms of country variations, although the [ESAC Registry of Transformative Agreements](https://esac-initiative.org/about/transformative-agreements/agreement-registry/) discloses an increasing number of similar national-level agreements, more than 80% of articles published in hybrid journals included in nationwide transformative agreements in Germany are still behind a paywall. Tracking open access across country affiliations reveals that uptake rates are particular low in the most productive countries USA, China and India. It remains to be seen whether a transition of hybrid journal portfolios to fully open access through transformative agreements is feasible given these observed global differences.

In future, we want to use hoaddata as data source for monitoring dashboards about the progress of nationwide transformative agreements in Germany, which we currently develop in the HOAD project with the support of the Deutsche Forschungsgemeinschaft. We want to extend its current scope on data about publication volume and open access uptake across agreements and countries. The aim is to highlight open metadata gaps not only relative to Creative Commons license information. Following up on our work on [metacheck](https://subugoe.github.io/metacheck/), an email tool to check metadata compliance, we will also present information about the coverage of Text and Data Mining support, funder infos, ORCIDs or Open Abstracts in Crossref metadata records; these data are critical for an open and reproducible monitoring of transformative agreements.

The launch of dashboards specific to nationwide transformative agreements in Germany is targeted for September 2022.

Borrego, Ángel, Lluís Anglada, and Ernest Abadal. 2020. “Transformative Agreements: Do They Pave the Way to Open Access?” *Learned Publishing* 34 (2): 216–32. <https://doi.org/10.1002/leap.1347>.

Marwick, Ben, Carl Boettiger, and Lincoln Mullen. 2018. “Packaging Data Analytical Work Reproducibly Using r (and Friends).” *The American Statistician* 72 (1): 80–88. <https://doi.org/10.1080/00031305.2017.1375986>.

Pollack, Philipp, Barbara Lindstrot, Irene Barbers, and Franziska Stanzel. 2022. “Open Access Monitor: Zeitschriftenlisten.” Jülich DATA. <https://doi.org/10.26165/JUELICH-DATA/VTQXLM>.

Schimmer, Ralf, Kai Geschuhn, and Andreas Vogler. 2015. “Disrupting the subscription journals’business model for the necessary large-scale transformation to open access.” Max Planck Digital Library. <https://doi.org/10.17617/1.3>.

1. <https://support.unpaywall.org/support/solutions/articles/44001894783-what-does-is-paratext-mean-in-the-api-> [↑](#footnote-ref-30)
2. See contract <https://pure.mpg.de/rest/items/item_3174351_1/component/file_3189424/content> [↑](#footnote-ref-44)