This is an Accepted Manuscript of an article published by Taylor & Francis in SERIALS REVIEW on 31 May 2022 available at: http://www.tandfonline.com/10.1080/00987913.2022.2066968.

Open Access Monitor Germany - Best practice in providing metrics for analysis and decision-making

Irene Barbers[a]* https://orcid.org/**0000-0003-2011-7444**

Franziska Stanzel[a] https://orcid.org/**0000-0003-0053-8604**

Bernhard Mittermaier[a] https://orcid.org/**0000-0002-3412-6168**

[a]*Central Library, Forschungszentrum Jülich GmbH, Jülich, Germany*

*i.barbers@fz-juelich.de

Open Access Monitor Germany - Best practice in providing metrics for analysis and decision-making

Abstract

In light of the ongoing transformation of scholarly publishing towards open access, libraries need data-based tools that enable them to make decisions that respond to the challenges posed by this shift. Tasks such as collection development, applications for funding, or consulting researchers on publication venues require a solid base of data, which is not always easily available to librarians. Funded by the German Federal Ministry of Education and Research (BMBF), the Open Access Monitor Germany (OAM) aims to provide libraries, funders, and researchers with a freely available tool that presents data on publications and citations for analysis. The OAM records the publication output of German academic institutions in scientific journals. Through analyses of subscription fees and (open access) publication fees, it helps to monitor and

support the transition of the publishing system towards an open access system. To this end, data from existing sources are collated within an expanded database, made available to users in an open interface, and disseminated by means of scientific publications. The OAM draws on sources such as Unpaywall, Dimensions, Web of Science, Scopus, and OpenAPC. Unpaywall is used as a central data source for publication metadata including open access availability; this data is then matched with Crossref data for journal-level metadata and finally with Dimensions, Web of Science and Scopus data for affiliations and citation data. The connection to OpenAPC provides data on publication costs for each participating institution. The interface https://open-access-monitor.de/publications presents users with a wide range of filters that can be used to customize the underlying data in line with specific needs. Users can switch between Dimensions, Web of Science and Scopus affiliations and make use of different analysis types including publication analyses, cost analyses, and citation analyses. Results are presented as tables as well as charts and can be downloaded for further use. In the future, additional data sources including subscriptions to scientific journals and subscription payments will be integrated to allow the shift in payment flows in the changing publication market to be observed. The integration of data from the electronic resource management system LAS:eR has already been initiated and will soon be completed. Moreover, interfaces will be created to connect the Alma and FOLIO systems. Together with data on publication costs, this will provide participating institutions with easily accessible overview of their total costs, thus enabling them to calculate an integrated budget for subscriptions and publications. Institutions using the OAM are advised with regard to analyzing data and technically integrating the OAM into their own applications. Small and medium-sized publishers from German-speaking countries are supported in the negotiation of transformation contracts by means of data

from the OAM. The OAM project team also conducts studies that address a number of scientific issues, for example the impact of open access on scientific networks and the role of open access in the field of monographs. The relationship between open access and the citation rates of published works is another possible research area that can be addressed with the help of the OAM.

Keywords

Scholarly publishing, scholarly communication, metrics, libraries, open access, data analysis

1. Introduction

"It's the data, stupid!" is a catchphrase attributed to Turing Award winner Jim Gray (Smith, 2009). The phrase is an adaptation of the slogan used during Bill Clinton's 1992 presidential campaign, but what does it mean? The long-standing scientific paradigms of experiment and theory were joined by a third paradigm—simulation—in the second half of the 20th century, and now a fourth paradigm has been added in the form of data science (Hey et al., 2009). The reason for this is that scientific instruments are producing more data than current methods can analyze, and so concepts and instruments must be developed to enable new knowledge to be gained from this flood of data. Now what does this have to do with libraries? Historically, libraries collected literature and made it available for use. Although methods for measuring this use have long been available (Butkovich, 1996), these only served to evaluate past decisions and subsequently adapt future decisions regarding subscriptions. The current transition of scientific publishing towards an open access system poses significant challenges for libraries—not only from an organizational and financial viewpoint, but also in terms of

the need for new ways of handling data.

In its conception, the Open Access Monitor Germany is probably unique in the world in the scope of the data it contains and its evaluation possibilities. The approaches in two other European countries are most comparable: The French Open Access Barometer [1] provides regular evaluations of the state of OA at institutional level using a methodologically similar approach to the Open Access Monitor, but is limited to publication data without including cost data. The Finnish portal JUULI [2] also does not include cost information, but offers very sophisticated search options. Like the Dutch solution NARCIS [3], it is based on an aggregation of repositories. Some of the Dutch repositories contain all publications, while others contain only open access publications. In Great Britain, a prototype was conceived (Johnson & Chiarell, 2017), but apparently not realized. The transnational approaches of the European Union [4] and the Curtin Open Knowledge Initiative COKI [5] should also be mentioned. Both, however, do not offer evaluation options on an institutional level and, like all the Open Access Monitors mentioned, do not contain cost information or publication data from multiple quality-checked data sources.

This paper will give a detailed analysis of why libraries have to work with data and what limitations they face in doing so. The Open Access Monitor (OAM) [6] will then be presented as a support tool for libraries in this work, and its practical benefits will be demonstrated based on a number of use cases.

1.1.Data handling

Following the launch of the Budapest Open Access Initiative (Chan et al., 2002), numerous science organizations from all continents signed the Berlin Declaration (Anonymous, 2003) to express their support for the transition of scientific publishing

towards an open access system. Since then, the process of transition has been initiated within scientific institutions and several publishers, and is in some cases at an encouragingly advanced stage; however, completion is still a long way off.

In Germany, there are presently two large transformative agreements in place. A majority of the German higher education and research institutions is subscribing to the Publish and Read contracts with Wiley and Springer Nature, negotiated by Project DEAL [7]. Similar agreements with a number of other publishers have been negotiated and are administered by nation-wide library consortia [8]. With the help of the first Open Access funding program started by the Deutsche Forschungsgemeinschaft (DFG) in 2010, many universities have been enabled to install a publication fund dedicated to cover at least part of their researchers' OA publications [9]. A second program has since been established by the DFG [10], and by contrast to the first one, not only higher education institutions but also non-educational public research institutions are now able to apply. While this new program continues to provide financial means for publications in Gold Open Access publications like its predecessor, Hybrid Open Access publications that are part of transformative agreements are now also eligible for funding.

In order to promote the transition taking place in this landscape across the board through the negotiation of and participation in open access or transformative agreements, information on publication output and on subscription and publication fees is required. Only within a few individual institutions, however, is this information available in its entirety.

- What journals do we subscribe to, and who publishes them?
- How often have the journals been used?

- How much did the journals cost?
- How many articles were published as open access and how many as closed access, and in which journals? Who are the journals published by?
- How often were these publications cited?
- What open access publication fees and other fees were paid for these publications?

The first three questions relate to traditional journal collection management and are therefore a standard part of library work. Most scientific libraries will be able to provide the answers. The last three questions relate to an institution's publications and clearly go beyond the traditional responsibilities of a library. No longer limited to managing third-party literature that the library makes available to its users, this work concerns publications from the library's institution. While these questions must be answered in order to consistently manage information on income/expenditure for scientific publications (Barbers et al., 2018), they frequently cause difficulty. In many cases, a complete bibliographic record for the institution is not available; in others, relevant databases have not been licensed or do not contain reliable information on corresponding authors. Information on expenditure for open access publication fees (Pieper & Broschinski, 2018) and especially for non-open access publication fees (Gray, 2015) is often not available centrally, as payment was not handled centrally (by the library). This is precisely where the Open Access Monitor (OAM) comes in, helping institutions to answer many of the questions being posed in relation to the transition.

In many cases, such organizational issues are outweighed by serious financial concerns. While the assumption is that the overall transition of the publishing system towards open access can be accomplished with the financial resources already available

(Schimmer et al., 2015), the implications for individual institutions vary considerably. Institutions that previously had high-volume contracts and/or that produce small numbers of publications (e.g., state libraries or institutions in the chemical industry) will tend to be in a better financial position in the transition to open access. Institutions with a low license volume, with particularly well-negotiated subscription agreements, and/or with a very high publication output (e.g., research-intensive institutions) will be in a comparatively worse position financially. An awareness of these correlations is important for individual institutions, but also for research funding bodies and—with regard to publicly funded institutions—for policymakers, who must ask the question: what adjustments need to be made to budgets and financial flows in an open access world?

---

[1] https://ministeresuprecherche.github.io/bso

[2] https://juuli.fi/

[3] https://www.narcis.nl/metrics/Language/en

[4] https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/trends-open-access-publications_en

[5] http://openknowledge.community/dashboards/coki-open-access-dashboard/

[6] Open Access Monitor: https://open-access-monitor.de/

[7] https://deal-operations.de/en/agreements

[8] https://esac-initiative.org/about/transformative-agreements/agreement-registry/

[9] https://www.dfg.de/en/dfg_profile/facts_figures/evaluation_studies_monitoring/studies/study_open_access/index.html

[10]
https://www.dfg.de/en/research_funding/programmes/infrastructure/lis/funding_opportunities/open_access_publication_funding/index.html

2. The OAM Concept

Monitoring the publication output of their researchers involves considerable effort on the part of individual institutions and cannot be achieved with traditional resources alone. Keeping track of output on a national level increases the effort involved many times over. To begin with, there are several possibilities for acquiring the necessary data. For instance, the data can be collected from the individual institutions and collated in a centralized manner. The institutions would have to have access to sufficient levels of the relevant data, which, as mentioned above, is frequently not the case. In addition, with this method, the majority of the data evaluation work must be carried out manually, which has a negative impact on reproducibility. A much more economic and sustainable method of data acquisition is to reuse existing sources as much as possible, applying consistent definitions and algorithms. The OAM acts as a central interface that brings together multiple source systems. This enables continuous monitoring at a national level and provides a basis for fact-based decisions and actions.

2.1. Source systems

The source systems used in the OAM can be seen in *Figure 1*, which shows that a multitude of resources are necessary to meet the total information requirements. One disadvantage of this is a dependence on the quality of the available data. It also requires mapping in order to produce complete data sets without duplicates. The relevant open access (OA) status or OA model are assigned based on Unpaywall [1] and the Directory

of Open Access Journals (DOAJ) [2]. Unpaywall is a database containing over 30 million OA scientific articles that are harvested from other databases such as PubMed. Content is also directly gathered from over 50,000 OA repositories. The key factor is that all articles harvested have been assigned a digital object identifier (DOI) (Dhakal, 2019). DOAJ contains over 17,000 fully OA journals, excluding hybrid models. In 2014, the quality criteria were adjusted in line with evolving framework conditions to ensure that DOAJ continues to function as a positive list and that "publish or perish" platforms are not indexed (Olijhoek et al., 2015).

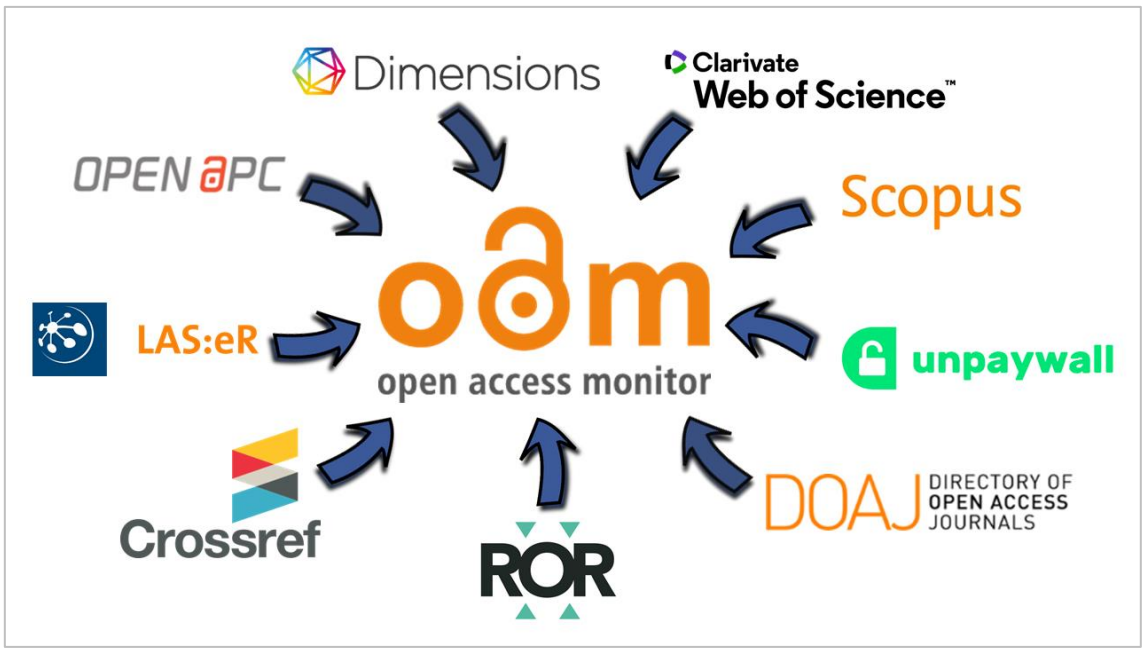


Figure 1: Overview of the source systems used in the OAM.

To ensure that the heterogeneous source data are sufficiently compatible for the matching process, they must be normalized. This requires a data set with a normalized format to allow the data from all source systems to be compared, harmonized, and then matched with the corresponding entries. Institutions were normalized on the basis of their GRID ID. The Global Research Identifier Database (GRID) [3] provides research institutions worldwide with a persistent identifier (PID) known as a GRID ID. This data resource provided the starting point for the community-based Research Organization

Registry (ROR) [4], which has been used in place of GRID for institutions in the OAM since the end of 2021 (ROR Leadership Team, 2021). Journals are normalized on the basis of Crossref [5]. Crossref enables cross-publisher linking between citing and cited publications by means of PIDs (Hendricks et al., 2020). It also provides a list of all journals registered with Crossref (Clark, 2020), which is used to perform normalization in the OAM.

Moreover, the data in the OAM have been supplemented from the beginning by data from Dimensions [6] and now also from Web of Science (WoS) [7] and Scopus [8]. Dimensions, like GRID, is run by Digital Science and offers a free version for personal use that contains fewer features than the paid version. Less selection is carried out than by competitor products, which leads to greater coverage of items (Schonfeld, 2018). In contrast, WoS uses a curation process that results in only a selection of journals being indexed in the citation database. Although Scopus also uses a selection process, the resulting coverage is somewhere between the levels provided by the other two databases. The OAM receives weekly updates from these three source systems via a data feed. While using three source systems to provide the same information may not seem an economic or efficient approach, all three databases have different data with differing strengths. Although Dimensions has the greatest coverage, it is less established than its competitor products thus far. WoS is advantageous for the OAM in that it indexes reprint or corresponding authors, while Scopus contains more titles from German-speaking countries (Singh et al., 2021). In addition, integrating all three databases in the OAM provides maximum data consistency with institutions' preferred data source, which reduces inconsistencies for users.

Gold and hybrid open access publications incur costs for the publishing institutions or authors, which must be taken into account as part of open access

monitoring. To do so, the data submitted by the institutions to OpenAPC [9] are used and supplemented with data from Crossref, PubMed, and DOAJ. (OpenAPC, 2021) The data resources are dependent on individual institutions submitting data as well as on the quality and completeness of that data. However, all too often libraries are not provided with sufficient information on the costs incurred for publications, in particular additional publication fees such as page charges and color charges. These costs can be incurred regardless of the publication model and are usually not documented centrally within institutions. The relevant data contained in the OAM are currently very patchy, as an adequate source system does not yet exist. Hopefully, this issue will be addressed by the openCOST project, and the required data will be added to OpenAPC for participating institutions in the future (Wagner, 2018); (Kloth, 2021). This will first involve creating the necessary infrastructure and developing a suitable metadata schema. The data can also be supplemented in the OAM by direct data submissions from institutions.

Subscription costs for journal licenses are harvested from LAS:eR [10]. LAS:eR is an electronic resource management system (ERMS) that allows consortial and local licenses as well as costing information to be managed in a single system. Institutions that use this system to record their license costs can share their data automatically with the OAM by providing the required consent. As agreements with publishers usually contain confidentiality clauses (Doenges, 2017), information regarding costs can only be made publicly available in aggregated form. To allow individual institutions to evaluate their total costs via the OAM nonetheless, a hierarchical rights management system is required to access this information. An authentication procedure is planned to ensure that subscription costs at an institutional level can only be viewed by the institution in question. While German institutions are not obliged to manage their

licenses via LAS:eR, and as it is not free of charge, the OAM is dependent on the use of LAS:eR by institutions, on their consenting to share their data with the OAM, and on the quality of the data they submit to LAS:eR. To reduce the OAM's dependency on the use of a particular ERMS and to offer more institutions the opportunity to monitor their subscription costs, the integration of additional library systems such as FOLIO and Alma is planned.

2.2. Workflow and backend

To generate a consistent data set from all of the source systems mentioned, Unpaywall is used as a starting point. The following data are collected for all journal articles via the weekly data feed: DOI, journal, publisher, publication date, and OA model. To assign OA categories to the articles, the checks listed in *Table 1* are run in the order shown. Once a condition has been fulfilled, the associated OA model is assigned and the checking process is terminated.

Table 1: Definition of open access (OA) models and internal assignment conditions.

OA model	Definition	Conditions
Diamond	Gold OA publication without article processing charges (APCs)	Gold OA without APCs
Gold	Published in an OA journal that is indexed in DOAJ	All articles in the journal are OA
Hybrid	Published in a subscription journal under an open license and freely accessible	Host is a publisher and article has a CC license
Green published	Accepted publisher's version	Host is a repository and article is the accepted publisher's version

Green accepted	Accepted manuscript version (postprint)	Host is a repository and article is the published version
Green submitted	Paid access on the publisher's website, but a free version of the submitted article is available in an OA repository (preprint)	Host is a repository
Bronze	Bronze: Free to read on the publisher's website, but no clearly recognizable license	Host is a publisher and article does not have a CC license
Closed	All other articles (not freely accessible)	Other

A comparison with an internal list is carried out to determine whether publishers are fully OA publishers. Journals' ISSNs are also compared with Crossref's list of journal titles. This step is required to reference journals and perform the subsequent mapping in order to achieve a consistent data pool with a single version of each journal title. On the basis of this list, the ISSNs are compared with the data from DOAJ to determine the journals' publication model. The journals are also compared with internal static journal lists. There is one static list for mirror journals, one for journals taken into account for funding applications to the German Research Foundation (DFG), and one each for journals included in individual transformative agreements. These lists are maintained by the OAM team and are publicly available [11]. An overview of the source systems used is provided in *Table 2*.

Table 2: OAM data sources and the data used from each one.

Purpose	Data source	Data used
Assignment of OA model	Unpaywall	Article data, OA model
	DOAJ	Journals' OA model

Normalization	Crossref	Journal normalization
	GRID/ROR	Institution normalization
Databases	Dimensions	Institutions, citation figures
	Web of Science	Institutions incl. corr. author, citation figures
	Scopus	Institutions incl. corr. author, citation figures
Costs	OpenAPC	Publication costs and article data
	LAS:eR	Subscription fees

Additional data are imported from Dimensions, WoS and Scopus via DOIs; for each article, the institution involved, database ID, and citation figures are added to the existing article data. For WoS and Scopus, the details of the corresponding author are also added. To achieve the greatest possible accuracy for assignments, the affiliated institution is assigned on the basis of its ROR ID. While Dimensions connects with ROR, WoS does not, which means that institution names from WoS need to be normalized to match the ROR IDs. Due to the strong heterogeneity of the data, this process is usually performed manually. A similar procedure is has been developed for institutions' names coming from Scopus. This data model allows OAM users to seamlessly switch between the different data sources.

To date, only information about the publication output of the individual institutions across Germany is available, with no details about the associated costs. To record the publication costs, an additional table is stored in the OAM backend. Data from OpenAPC on the individual articles, such as the DOI, publication year, and the institution paying the fees, are taken and likewise mapped to the ROR ID. The costs are also recorded, with a distinction made between APCs and hybrid articles. These data are

supplemented with additional data from Unpaywall, with assignment taking place on the basis of the DOI. The journal's publisher, publication date, journal, and OA model are recorded.

While this means that the database also contains the costs for OA publications, subscription fees are not yet included. These have not yet been implemented in the production system; however, the integration of LAS:eR in the test system has already been completed. In accordance with the EU General Data Protection Regulation, each institution must consent to the sharing of their data in LAS:er; many institutions have not yet done so. In addition, an authentication mechanism will be set up to allow institutions to access their data without it being viewed by third parties. A number of options are currently being evaluated in consultation with LAS:eR.

2.3. Frontend

The frontend of the OAM is implemented via the open-source development platform ASP.NET [12], using Blazor [13] as a framework. MudBlazor [14] provides the framework for applications and Plotly [15] supports a wide range of analysis options. The user interface is currently bilingual English / German. The language on display is adapted to the web browser settings and can be switched in the upper right-hand corner.

An overview of German OA publications (*Figure 2*) is presented on the start page of the application [16]. A graph shows the distribution of APCs, while a separate chart provides a breakdown of the articles according to publication model.

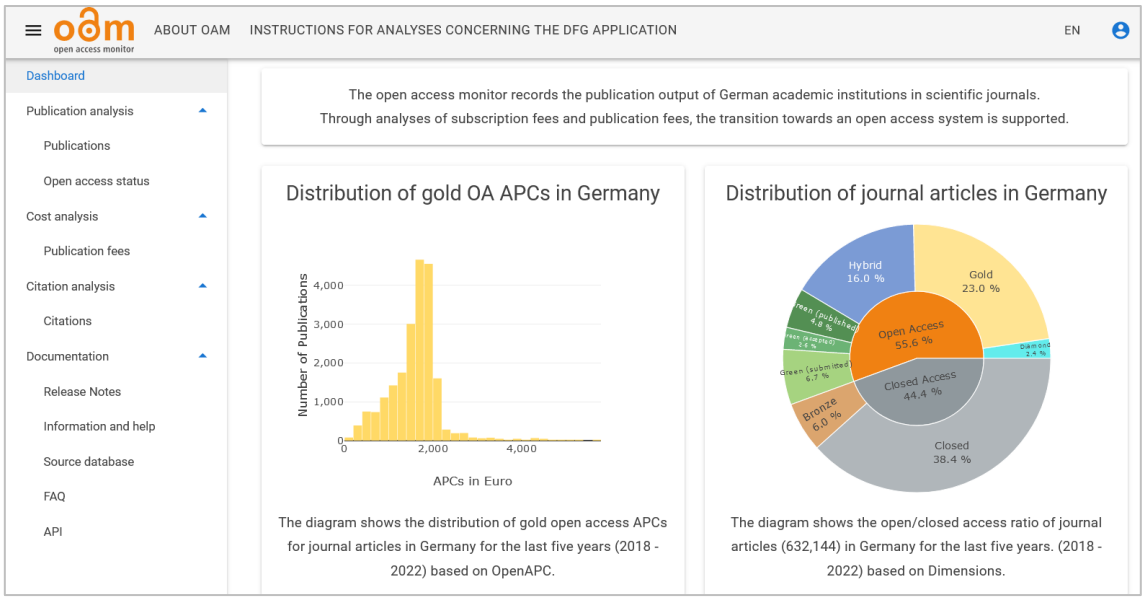


Figure 2: The OAM start page with two diagrams showing the distribution of APCs for gold OA publications in Germany and the distribution of OA models for the published articles.

The menu on the left-hand side of the page allows users to navigate to different areas of the website: Publication analysis, Cost analysis, Citation analysis, and Documentation. The items at the bottom of the navigation bar relate to documentation and link to the OAM Wiki pages, which contain additional information on the source databases, the API, and FAQs. Selecting “Publications” under “Publication analysis” in the menu displays a page with numerous filter options, as shown in *Figure 3*.

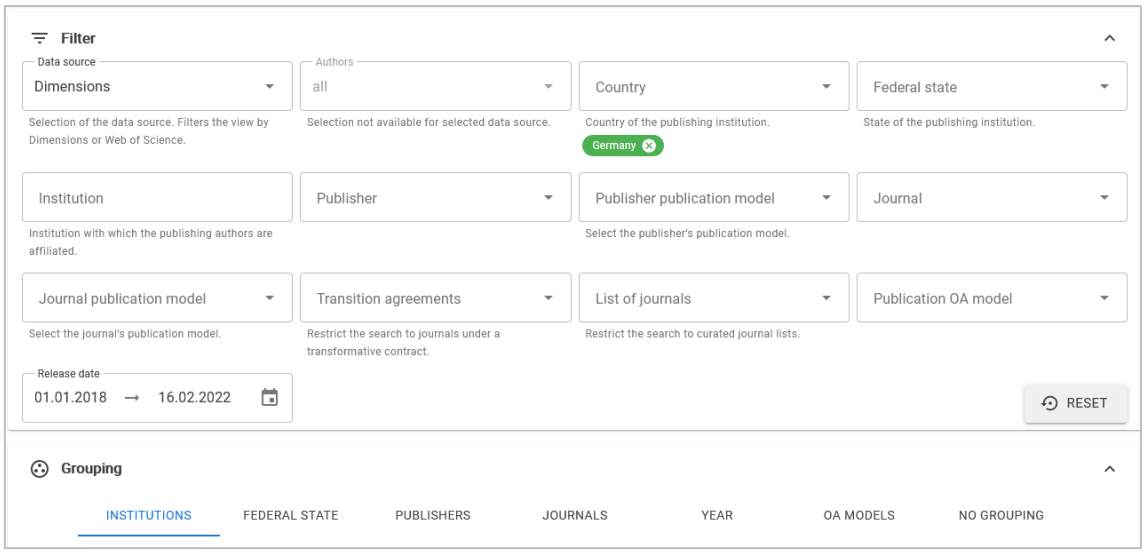


Figure 3: Filters for publication analyses and grouping options.

These filters allow users to switch between data sources for affiliations, select the country, and—for WoS and Scopus—restrict results to corresponding author publications. Filtering by federal state, institution, publisher, journal, and OA model is also possible. The names of each have been normalized. When filtering by OA model, users can select the OA model of the publisher, the journal, or the individual article. In addition, the "Transition agreements" filter makes it possible to select journals covered by individual national transformative agreements in Germany. Publications can also be filtered by publication date, which supports analyses of periods within a given year. Grouping options are also available, which allow users to display the selected database entries according to their specific needs. Entries can be grouped by institution, federal state, publisher, journal, year, and OA model, and it is also possible to view the publications including their DOI without grouping.

In addition to listing entries in tabular form, the OAM offers three options for graphic representation: bar charts, line charts, and pie charts. Users can download these graphs as PNG files and the data from the tables displayed above as CSV files in order to carry out their own analyses.

A sharing function is provided to facilitate communication on analyses. The shared link contains the previously selected filter and grouping settings as well as the chart selection, and displays up-to-date analysis results.

---

[1] Unpaywall: http://unpaywall.org/

[2] Directory of Open Access Journals: https://doaj.org/

[3] Global Research Identifier Database: https://www.grid.ac/

[4] Research Organization Registry: https://ror.org/

[5] Crossref: https://www.crossref.org/

[6] Digital Science: Dimensions. https://www.dimensions.ai/

[7] Clarivate: Web of Science. https://clarivate.com/webofsciencegroup/solutions/web-of-science/

[8] Elsevier: Scopus. https://www.elsevier.com/solutions/scopus

[9] OpenAPC: https://openapc.net/

[10] University library center of the Federal State of North Rhine-Westphalia (hbz): LAS:eR. https://www.hbz-nrw.de/produkte/digitale-inhalte/las-er

[11] Open Access Monitor: Journal lists. https://doi.org/10.26165/JUELICH-DATA/VTQXLM.

[12] ASP.NET: https://dotnet.microsoft.com/apps/aspnet

[13] Blazor: https://dotnet.microsoft.com/apps/aspnet/web-apps/blazor

[14] MudBlazor: https://mudblazor.com/

[15] Plotly: https://plotly.com/

[16] Open Access Monitor: https://open-access-monitor.de/

3. Use Cases

The services provided by the OAM are tailored to the needs of various stakeholders. For instance, libraries as well as consortial or interregional negotiation teams need to evaluate or negotiate transformative agreements based on concrete data, and plan for reallocations between their acquisition and publication budgets. Meanwhile, institutions' needs include monitoring their own publication output, identifying the areas that publish open access less frequently, and promoting open access in a targeted manner. They also need to analyze data in relation to funding applications. The OAM provides a diverse range of data as well as analyses to meet all these needs. In addition, scientific studies carried out as part of the OAM project provide insights into the current state of the open access transition in Germany and internationally (Mittermaier, 2021); (Barbers & Pollack, 2021) and contribute to the further development of data-based

monitoring methods.

3.1. Publication analyses

Individual institutions can use the OAM's services for many different purposes. For instance, the OAM data can be used to perform comparisons with institutions' internal publication databases. Creating publication lists by performing an analysis of all publications from a particular institution within a given period and downloading the results is also possible. These lists can then be used to check and, where necessary, to update institutions' own publication databases (bibliography).

An institution's publication output in relation to the evolution of its open access status can be queried by selecting the relevant institution using the "Institution" filter. The results can help to assess the development of the publication behavior of the institution's researchers and the impact of budget reallocations from subscription fees to publication fees. In *Figure 4*, for example, the publication output of Forschungszentrum Jülich is displayed by means of a bar chart depicting the total number of publications. The grouping settings make it possible to display the publications in different colors according to their open access categories, which allows the development of individual categories such as gold open access to be easily identified and compared.

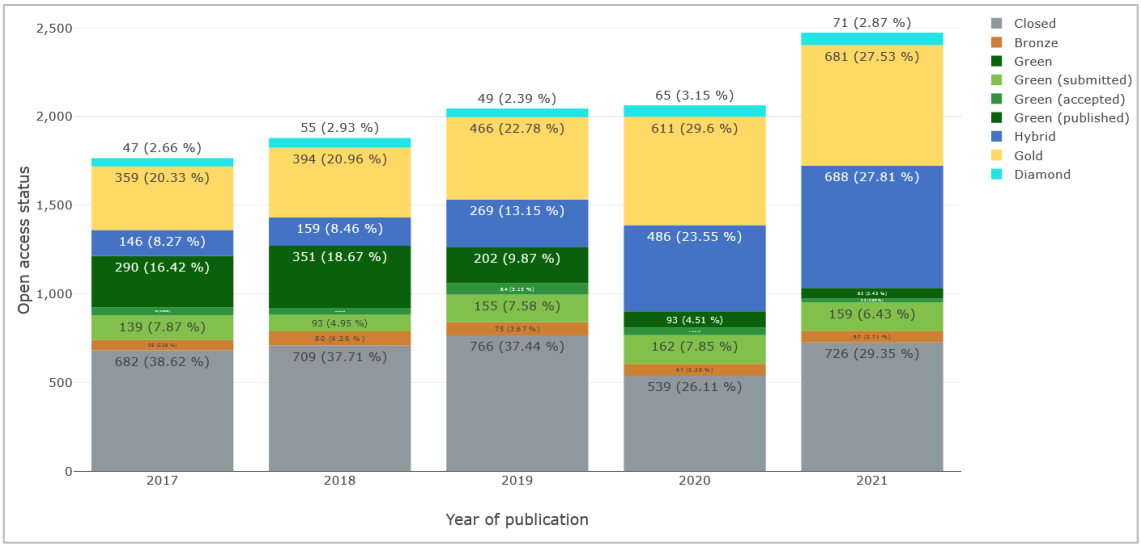


Figure 4: Graphical representation of the publications of Forschungszentrum Jülich from 2017 to 2021, broken down according to open access categories. Source data: Unpaywall/Dimensions.

3.2. Data to support funding applications

The announcement of the DFG's new Open Access Publication Funding program [1] opened up a new field of action for the OAM's services. As part of the funding application, institutions must submit comprehensive data analyses of their own publication output in previous years in order to create forecasts for the number of articles to be funded. As many institutions do not yet have adequate infrastructures or processes in place to record data on their own publication output, the OAM offers support to those institutions in performing the necessary data analyses. By providing a specific filter for the journal portfolios covered by transformative agreements and a curated list of open access journals that meet the DFG's funding criteria ("quality-assured open access publication channels"), the OAM enables institutions to efficiently gather the data required for their applications. A set of instructions has been made available to explain the steps and filter settings required in the user interface [2]. The journal lists have been published in the Jülich DATA repository [3].

Institutions that wish to submit funding applications can retrieve information on the publication output of individual authors, filtered by corresponding author publications, in just a few steps. The analysis should include both publications published in fully OA journals and publications from transformative agreements. The filters should be used to select the institution applying for funding as well as all relevant transformative agreements, as shown in *Figure 5*. A period of time corresponding to the three publication years preceding the application year should be selected, and the search should be restricted to corresponding authors.

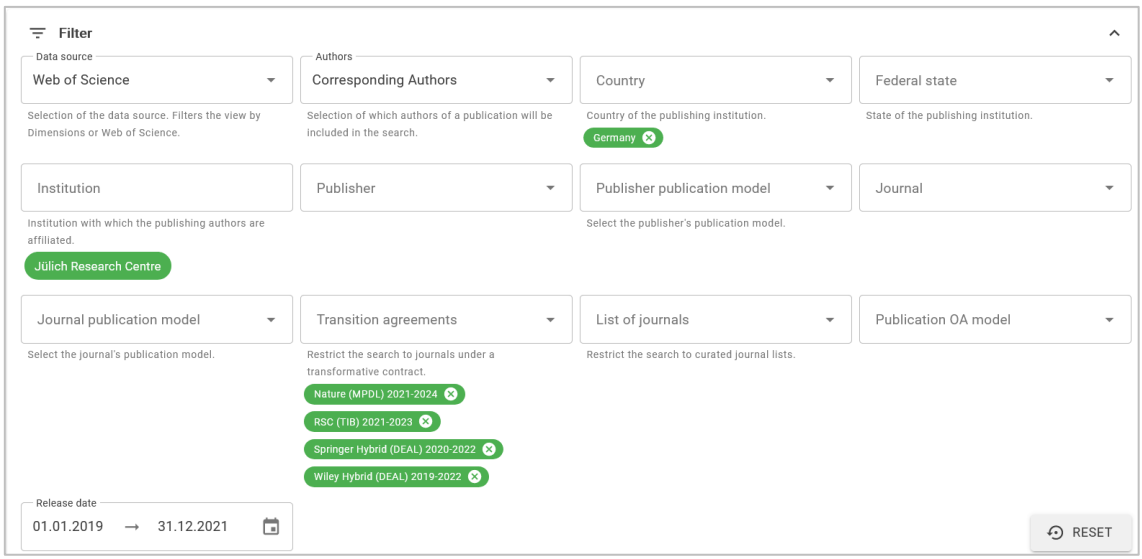


Figure 5: OAM filters required to analyze publication output in transformative agreements: institution, transformative agreements, corresponding authors, and publication date.

3.3. Performing analyses to support negotiations

Negotiations on transformative agreements aim to transfer existing or newly established consortia from subscription-based agreements to agreements that provide the participating institutions with both reading access and open access publishing services, ideally without any additional fees. The preparatory work for such negotiations requires intensive data-based analyses to enable the negotiation team to assess the impact on

individual institutions and on the consortium as a whole. The OAM can help teams to answer questions such as: How many publications by authors from German institutions were published in the last few years by a particular publisher? How has the situation developed in terms of the occurrence of the different open access types? While *Figure 6* and *Figure 7* provide an initial overview comparing the developments at two publishers, it is also possible to group the results by institutions and to thus obtain a detailed insight into the publication figures of individual institutions, and hence of potential parties to an agreement (*Figure 8* and *Figure 9*).

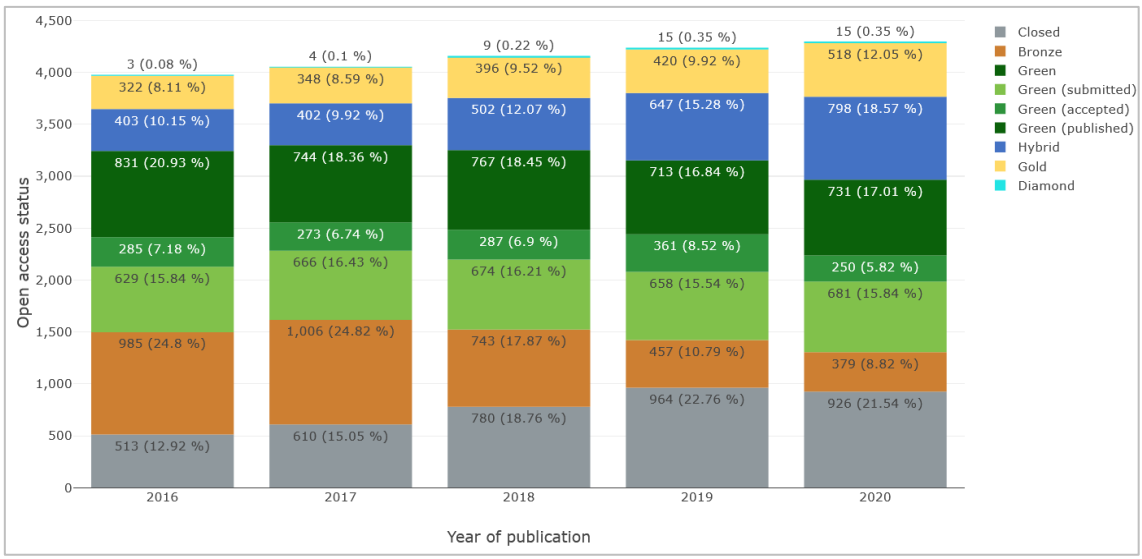


Figure 6: Number of German publications by Oxford University Press from 2016 to 2020, broken down by open access status. Source data: Unpaywall/Web of Science.

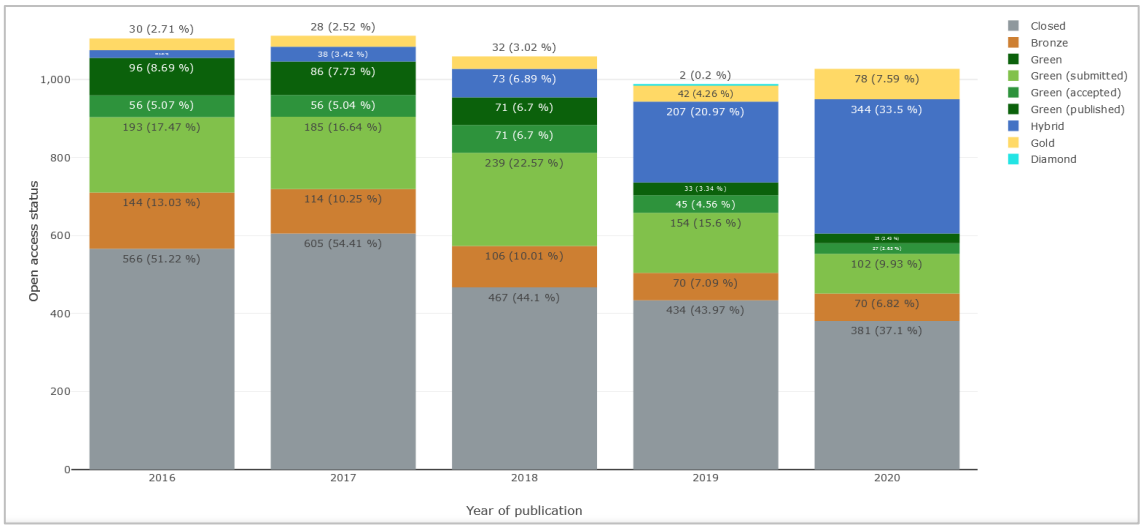


Figure 7: Number of German publications by Cambridge University Press from 2016 to 2020, broken down by open access status. Source data: Unpaywall/Web of Science.

The publications by Oxford University Press (OUP) slightly increase over time (*Figure 6*) in line with overall publication trends, whereas the figures from Cambridge University Press (CUP) show a decline (*Figure 7*). The increase in the proportion of hybrid OA publications by CUP is particularly striking, which can be attributed to the transformative agreement concluded between Bayerische Staatsbibliothek and CUP. A corresponding transformative agreement has not been concluded with OUP. This also highlights the fact that while transformative agreements may lead to an increase in the proportion of OA articles published, they do not necessarily result in an increase in publication figures (Mittermaier, 2021).

Breakdowns of publication figures by institution also provide an insight into the varying perceptions of secondary publication rights and the publication of preprint versions, both among the different publishers and among institutions. The following examples show the publication output of the ten German institutions with the highest publication figures at OUP (*Figure 8*) and CUP (*Figure 9*). The extremely high proportion of green OA for the three Max Planck Institutes and the European Southern Observatory is worth noting (*Figure 8*). This is due to two reasons: firstly, the liberal

green OA policy (no embargo period) applied by the *Monthly Notices of the Royal Astronomical Society* journal, in which these institutions publish a great deal; secondly, the willingness of their researchers to make use of the options available. In many institutions, the libraries systematically enter publications in repositories on behalf of the authors. In the case of CUP there is no subject-related specialty, which means that the institutions are easier to compare. The proportion of green OA articles by LMU Munich and, in particular, HU Berlin stands out.

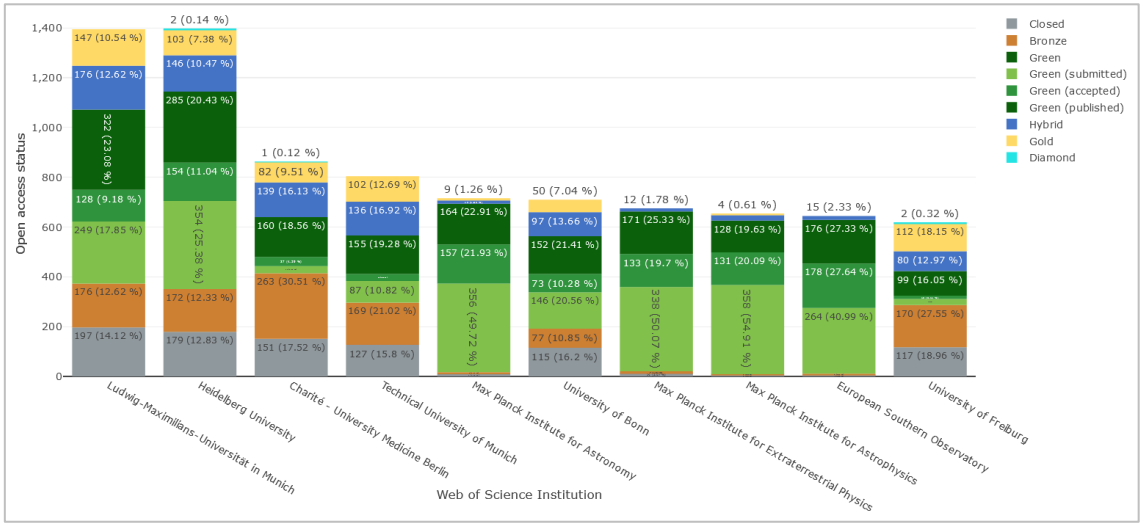


Figure 8: Output of the ten German institutions with the highest publication figures at Oxford University Press from 2016 to 2020, broken down by open access status. Source data: Unpaywall/Web of Science.

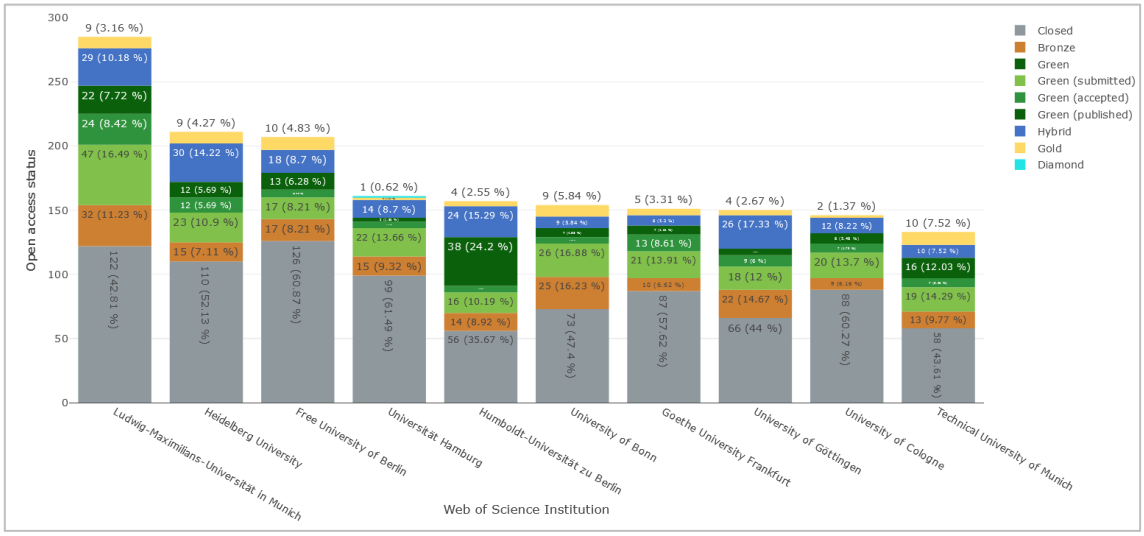


Figure 9: Output of the ten German institutions with the highest publication figures at Cambridge University Press from 2016 to 2020, broken down by open access status. Source data: Unpaywall/Web of Science.

3.4. Cost analyses

Due to the integration with OpenAPC, the functions already implemented in the OAM allow gold OA and hybrid OA publication fees to be displayed for institutions that submit information on their expenses to OpenAPC. The OAM interface offers more grouping and representation options than OpenAPC. In particular, in addition to the fees mentioned above, other charges can be identified—the "hidden" or "traditional" publication fees that in some cases can substantially increase institutions' expenditure. Data on color charges, page charges, etc., can therefore be supplied by the institutions directly to the OAM for inclusion in the analyses. During the further course of the project, integrations with ERMSs will also deliver information on licenses and subscription fees to the OAM and complete the picture both on the level of individual institutions and, ultimately, on a national level. The OAM will thus be able to reach its full potential as a source of support within the open access transition. Ideally, institutions will in the future receive a complete overview of their portfolio of licenses, publications, and the associated fees to support them in managing information on their income/expenditure for scientific publications. The goal is to provide each participating institution with data via the Jülich Open Access Barometer [4], which currently exists on a local level and is shown in *Figure 10*.

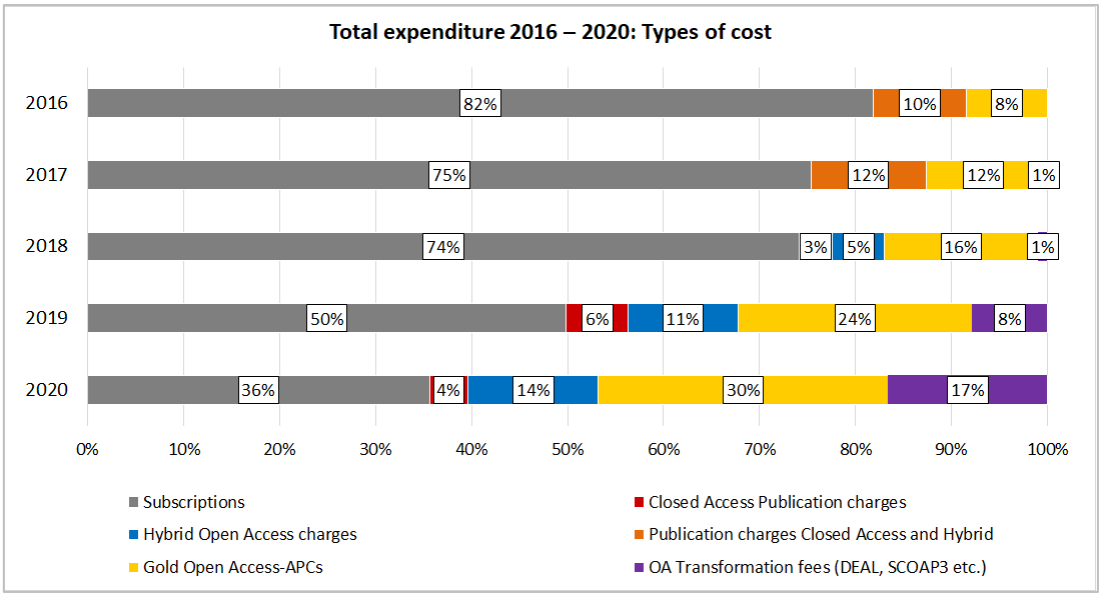


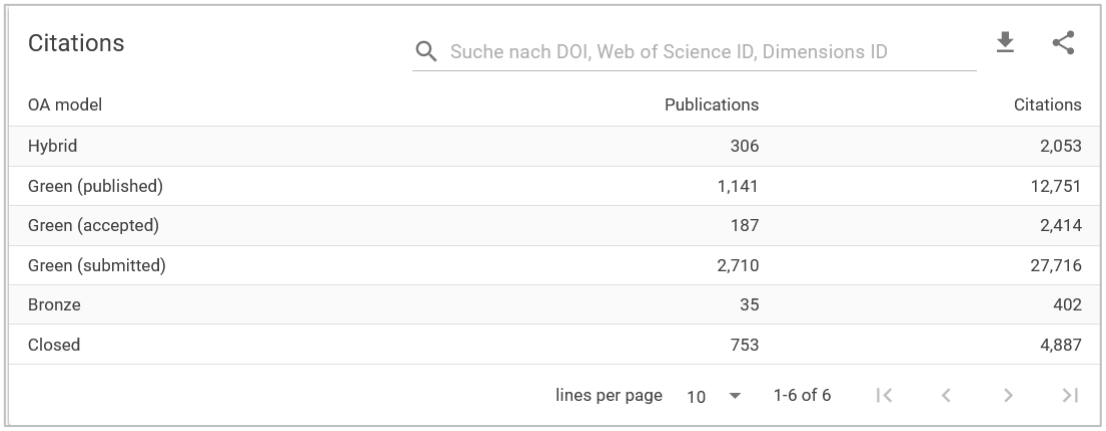
Figure 10: Jülich Open Access Barometer showing the share of expend Jülich.

3.5. Citation analyses

In connection with the OA status, citation analyses can provide insights into how the publications of a journal or a publisher are received. Although the representations produced by the OAM do not provide information about the citation behavior of individual authors, they can for example indicate how frequently the publications of authors from a particular institution were cited at journal, publisher, regional, or national level. The results can also be filtered, for example to show transformative agreements.

OAM analyses can also be used to support investigations of any citation advantage that may result from OA publishing. For instance, *Figure 11* shows the number of publications in the journal *Physical Review B* broken down by open access types, including the number of citations for each type. Even though the citation rate can be significantly influenced by individual publications, the example below shows that freely accessible publications are cited much more frequently than those published

closed access. Green OA publications have an average citation rate of 9.0, with hybrid OA publications reaching a rate of 11.9. With a citation rate of 5.9, closed OA publications are cited only half as frequently.



Citations

Suche nach DOI, Web of Science ID, Dimensions ID

OA model	Publications	Citations
Hybrid	306	2,053
Green (published)	1,141	12,751
Green (accepted)	187	2,414
Green (submitted)	2,710	27,716
Bronze	35	402
Closed	753	4,887

lines per page 10 1-6 of 6

Figure 11: Distribution of citations for publications involving German authors in the journal Physical Review B from 2017 to 2021, broken down by the open access models of the cited publications. Source data: Unpaywall/Dimensions.

---

[1] DFG Open Access Publication Funding program: (https://www.dfg.de/en/research_funding/programmes/infrastructure/lis/funding_opportunities/open_access_publication_funding/index.html

[2] The instructions are available in German here: http://hdl.handle.net/2128/26338.

[3] OAM journal lists: https://doi.org/10.26165/JUELICH-DATA/VTQXLM.

[4] Jülich Open Access Barometer: https://www.fz-juelich.de/zb/EN/Expertise/open_access/oa_barometer/_node.html.

4. Conclusions

The Open Access Monitor Germany (OAM) has established itself as an intensively used tool for scientific institutions and libraries. It has already provided important services for the open access transition at a national level, and continues to do so. The OAM offers support for data handling and provides metrics for decision-making as part of collection management. The core long-term task of the OAM infrastructure is to

systematically combine the data from the various sources and thus to continuously normalize journals and publishers in addition to verifying the assignment of affiliations.

The use of databases like Dimensions, Web of Science and Scopus as sources has the caveat of dependency on continuing agreements and services, however these databases have the advantage of being regarded as authoritative and are renowned within the community. Moreover, Web of Science and Scopus are the only known sources that provide corresponding author information for articles, which is essential for institutions to make a reliable prognosis for their budget planning. Nonetheless, the OAM will look out for emerging new data sources, especially open sources, and will evaluate options for integrating their data.

The underlying work of normalizing source data for institutional affiliations cannot be underestimated. Even with the use of algorithms looking for similar strings to match lists of names, there is a large amount of manual work left to do; and a good knowledge of the research landscape is requisite, as well as resourcefulness and search skills. The use of ROR as an institutional identifier in source databases can therefore be considered as a huge benefit for the integration process.

During the further course of the project, new functions will be created and additional data sources integrated to improve and expand the spectrum of analysis. The integration of library systems such as FOLIO and Alma is planned in order to provide comprehensive monitoring services to a greater number of libraries. Finally, the integration of usage statistics through a link with the German national statistics server (Wahlig, 2018) will enable users to obtain all of the data required for collection management in one place.

The OAM's services are also available internationally. On behalf of the Consortium of Swiss Academic Libraries (CSAL)[1], an application is currently in

development to provide national open access monitoring for Switzerland and make the service available via a user interface, similar to the German version.

The OAM project, which is being funded by the German Federal Ministry of Education and Research (BMBF) in its second phase, will be put into regular operation after the funding expires in 2023, from which point it will be funded by Forschungszentrum Jülich's own resources. The aim is to provide the user community with a sustainable, freely accessible tool that is continually adapted to users' requirements.

---

[1] Consortium of Swiss Academic Libraries (CSAL): https://consortium.ch/?lang=en

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF) under Grant FKZ 16OAMO001

Declaration of Interest Statement

The authors report there are no competing interests to declare.

References

Anonymous. (2003). *Berlin declaration on open access to knowledge in the sciences and humanities*. https://openaccess.mpg.de/67605/berlin_declaration_engl.pdf

Barbers, I., Kalinna, N., & Mittermaier, B. (2018). Data-driven transition: Joint reporting of subscription expenditure and publication costs. *Publications*, *6*(2), 19. https://doi.org/10.3390/publications6020019

Barbers, I., & Pollack, P. (2021). *Open Access in Deutschland: Entwicklung in den Jahren 2005 - 2019*. Forschungszentrum Jülich, Zentralbibliothek. http://hdl.handle.net/2128/27849

Butkovich, N. J. (1996). Use studies: A selective review. *Library Resources & Technical Services*, *40*, 359–368

Chan, L., Cuplinskas, D., Eisen, M., Friend, F., Genova, Y., Guédon, J.-C., Hagemann, M., Harnad, S., Johnson, R., Kupryte, R., LaManfredi, Rév, I., Segbert, M., deSidnei, Suber, P., & Velterop, J. (2002). *Budapest open access initiative*. https://www.budapestopenaccessinitiative.org/read

Clark, R. (2020). *Browsable title list*. https://www.crossref.org/documentation/reports/browsable-title-list/

Dhakal, K. (2019). Unpaywall. *Journal of the Medical Library Association: JMLA*, *107*(2), 286–288. https://doi.org/10.5195/jmla.2019.650

Doenges, H. (2017). *Non-disclosure clauses: The making, breaking, and remaking of relationships* [PhD thesis, University of Washington; School of Law]. https://depts.washington.edu/uwlawlib/wordpress/wp-content/uploads/2018/01/Doenges2017.pdf

Gray, A. (2015). Considering non-open access publication charges in the "total cost of publication." *Publications*, *3*(4), 248–262. https://doi.org/10.3390/publications3040248

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, *1*(1), 414–427. https://doi.org/10.1162/qss_a_00022

Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research. https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/

Johnson, Rob & Chiarell, Andrea (2017). Defining and Prototyping an Open Access Dashboard. Final Report, Prepared on Behalf of JISC. https://scholarlycommunications.jiscinvolve.org/wp/files/2017/09/Defining-and-Prototyping-an-OA-Dashboard-Final-Report-Abridged.pdf

Kloth, H. (2021). *openCost - Universität Regensburg*. https://www.uni-regensburg.de/bibliothek/projekte/opencost/index.html

Mittermaier, B. (2021). Transformationsverträge – Stairway to Heaven oder Highway to Hell? *027.7 Zeitschrift Für Bibliothekskultur / Journal for Library Culture*, *8*(2). https://doi.org/10.21428/1bfadeb6.d80f0652

Olijhoek, T., Bjørnshauge, L., & Mitchell, D. (2015). Criteria for open access and publishing. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AMHUHV.v1

OpenAPC. (2021). *Collect and disseminate information on fee-based Open Access publishing.* https://github.com/OpenAPC/openapc-de

Pieper, D., & Broschinski, C. (2018). OpenAPC: A contribution to a transparent and reproducible monitoring of fee-based open access publishing across institutions and nations. *Insights*, *31*. https://doi.org/10.1629/uksg.439

ROR Leadership Team. (2021). *ROR and GRID: The way forward*. https://ror.org/blog/2021-07-12-ror-grid-the-way-forward/

Schimmer, R., Geschuhn, K. K., & Vogler, A. (2015). *Disrupting the subscription journals' business model for the necessary large-scale transformation to open access.* https://doi.org/10.17617/1.3

Schonfeld, R. C. (2018). *A new citation database launches today: dimensions*. https://scholarlykitchen.sspnet.org/2018/01/15/new-citation-database-dimensions/

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of web of science, scopus and dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

Smith, D. (2009). *According to Microsoft, the fourth paradigm of science is data* (Revolution Analytics, Ed.). https://blog.revolutionanalytics.com/2009/12/fourth-paradigm.html

Wagner, A. (2018). *APC-Verwaltung im institutionellen Repositorium: GMS Medizin - Bibliothek - Information; 18(3):Doc21*. https://doi.org/10.3205/MBI000422

Wahlig, R. (2018). *Der Nationale Statistikserver. Vom Projekt zum nationalen Dienst*. https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/3583