Scientometrics (2021) 126:9751–9777
https://doi.org/10.1007/s11192-021-04002-0

Open access uptake in Germany 2010–2018: adoption in a diverse research landscape

Anne Hobert[1] · **Najko Jahn**[1] · **Philipp Mayr**[2,3] · **Birgit Schmidt**[1] · **Niels Taubert**[4]

Received: 15 June 2020 / Accepted: 17 April 2021 / Published online: 23 May 2021

Abstract

This study investigates the development of open access (OA) to journal articles from authors affiliated with German universities and non-university research institutions in the period 2010–2018. Beyond determining the overall share of openly available articles, a systematic classification of distinct categories of OA publishing allowed us to identify different patterns of adoption of OA. Taking into account the particularities of the German research landscape, variations in terms of productivity, OA uptake and approaches to OA are examined at the meso-level and possible explanations are discussed. The development of the OA uptake is analysed for the different research sectors in Germany (universities, non-university research institutes of the Helmholtz Association, Fraunhofer Society, Max Planck Society, Leibniz Association, and government research agencies). Combining several data sources (incl. Web of Science, Unpaywall, an authority file of standardised German affiliation information, the ISSN-Gold-OA 3.0 list, and OpenDOAR), the study confirms the growth of the OA share mirroring the international trend reported in related studies. We found that 45% of all considered articles during the observed period were openly available at the time of analysis. Our findings show that subject-specific repositories are the most prevalent type of OA. However, the percentages for publication in fully OA journals and OA via institutional repositories show similarly steep increases. Enabling data-driven decision-making regarding the implementation of OA in Germany at the institutional level, the results of this study furthermore can serve as a baseline to assess the impact recent transformative agreements with major publishers will likely have on scholarly communication.

Keywords Open access publishing · Scholarly communication · Empirical study · German research landscape · Open access classification

✉ Anne Hobert
hobert@sub.uni-goettingen.de

[1] Göttingen State and University Library, University of Göttingen, Göttingen, Germany

[2] GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany

[3] Institute of Computer Science, University of Göttingen, Göttingen, Germany

[4] Institute for Interdisciplinary Studies of Science (I2SoS), Bielefeld University, Bielefeld, Germany

Introduction

Open access (OA) to scholarly literature, defined as being "digital, online, free of charge, and free of most copyright and licensing restrictions" (Suber, 2012), has gained a prominent position on the agendas of research, policy-makers and academic publishing. Consequently, an evolving body of bibliometric research investigates the uptake of this publishing model. While strong evidence exists that OA is growing (Archambault et al., 2014; Laakso & Björk, 2012; Piwowar et al., 2018), quantitative studies focusing on countries and research institutions revealed notable variations from the general trend (Bosman & Kramer, 2018; Martin-Martin et al., 2018; Huang, 2020; Robinson-Garcia et al., 2020, Pölönen et al., 2020). Here, we contribute to this evolving evidence-base of OA to journal articles at the level of institutions with an analysis of the situation in Germany between 2010 and 2018.

Investigating the German research landscape and its OA uptake is of particular interest for several reasons: The first is Germany's country-specific broad range of mainly publicly funded universities and non-university research organisations. Germany is not only a country with a strong publication output in terms of journal articles (Wohlgemuth et al., 2017; National Science Board, National Science Foundation, 2019; Stahlschmidt et al., 2019), but it also has a diverse landscape of research institutions producing these research outputs. In addition to universities, significant parts of basic and applied research is conducted by non-university research institutions belonging to other sectors, each of them having different functions and different missions in the German research landscape.[1] Yet, institutional OA studies and rankings often restrict their focus on universities (e.g., Abediyarandi & Mayr, 2019), overlooking the impact of non-university research in Germany and other countries (Rovira et al., 2019). An analysis at the level of institutions by sectors that takes into account their different functions therefore complements investigations related to the OA uptake of universities.

Another reason is that German research institutions and organisations are not just early adopters of OA, but have also shaped European and global OA policies. Prominent examples are the Berlin Declaration on Open Access[2] (2003) and the recent OA 2020 Initiative,[3] calling for transitioning subscription-based journal publishing to OA. Since then, Germany's large diversity of institutionalised forms of research organisations has led to a decentralised adoption of OA. By contrast, some other European countries followed a more centralised approach coordinated by national research funding bodies. For instance, the United Kingdom, a country with similar research productivity in terms of journal publications, implements national OA policies with a strong focus on providing fee-based OA via journals and centralised management of OA funding streams via block grants (Pinfield et al., 2016). Similar to the international situation, German funders and research organisations alike increasingly negotiate transformative agreements, in which spending for subscriptions and open access publications are considered together, focusing on large publishers at the national level. In particular, the broad cancelation of Elsevier journals caused international attention (Else, 2018). However, the DEAL consortium,[4] comprising most universities and non-university research institutions in Germany, has successfully negotiated

[1] See section Public Research Landscape in Germany.
[2] https://openaccess.mpg.de/Berlin-Declaration.
[3] https://oa2020.org/.
[4] https://www.projekt-deal.de/.

agreements with Wiley and Springer Nature that came into effect in 2019, and in the beginning of 2020, respectively (Vogel, 2019).

Previous bibliometric studies devoted to OA at the institutional level complement global (Archambault et al., 2014; Laakso & Björk, 2012; Piwowar et al., 2018), disciplinary (Severin et al., 2020) and funder-specific analyses (Larivière & Sugimoto, 2018). In particular, Huang et al. (2020) argue for institutional OA studies because of policy interventions. The authors investigated the influence of external policies and funder requirements on the OA publication output of universities. They found varying OA uptake levels and differing OA adoption strategies across universities and countries. In 2019, the CWTS Leiden Ranking[5] started to present OA indicators. Analysing the underlying evidence-base revealed notable discrepancies between countries and institutions in terms of the OA adoption (Robinson-Garcia et al., 2020). Both related studies stressed conceptual and methodological challenges. Most importantly, affiliation information from bibliometric databases needed to be cleaned and normalised before carrying out the institutional level analyses. Likewise, not only choices of bibliographic data sources were critical, but also a data-driven classification of OA types. While the OA discovery service Unpaywall[6] has become the de facto standard for OA bibliometric studies, the studies argued that additional OA evidence sources need to be integrated.

Against this background, our study has three main objectives. First, we aim to determine the extent to which the journal publication output of German research institutions covered by the Web of Science is OA. The point in time of the analysis is chosen strategically. It focuses on the publication period from 2010 until 2018 and was conducted before a widespread adoption of transformation agreements. As large impacts of these agreements can be expected (Schimmer et al., 2015), the results reported in this article can serve as a baseline for future studies.

Second, a bibliometric analysis of the German research system needs to reflect its complexity. We will therefore review the different sectors and their specific missions, and how they are shaped by a discipline or subject field. For the bibliometric investigation, we draw on disambiguated affiliation information for German universities and non-university research organisations included in the Web of Science in-house database from the Competence Center for Bibliometrics[7] (Rimmert et al., 2017). This allows determining institutional publication activities, as well as to compare them against their specific missions and disciplinary profiles.

Finally, the paper does not stop at reporting the overall OA uptake in Germany, but also highlights different OA adoption strategies. For this aim, a data-driven classification effort combines data from the widely used OA discovery service Unpaywall with journal-specific open access status information from the ISSN-GOLD-OA 3.0 list (Bruns et al., 2019) and repository metadata from the Directory of Open Access Repositories (OpenDOAR). This allows us to extend and further describe which OA patterns were most commonly adopted in Germany.

Following this approach, the article addresses the following research questions:

RQ1: How did the OA fraction of the publication output of German universities and non-university research institutions develop over the period 2010–2018?

[5] https://www.leidenranking.com/.

[6] https://unpaywall.org/.

[7] http://www.forschungsinfo.de/Bibliometrie/en/.

RQ2: Which differences between the research sectors of the German research system can be found in terms of OA adoption and what are possible explanations for them?
RQ3: Which OA approach is most prevalent, and is it possible to identify different patterns of adoption to OA?

The next section will review the research landscape and OA adoption in Germany. After that, we will present our OA classification and describe how we obtained our data. Results are presented and discussed for each research question, and followed by general conclusions.

Background

Public research landscape in Germany

Compared with other countries, the German research system is comprised of a large diversity of different types of organisations (Powell & Dusdal, 2017). The research system consists of a private sector (i.e., research units funded by for-profit companies) and a public sector (i.e., research organisations that receive basic funding from the German government, one of Germany's 16 federal states, or a combination of both). The public sector is differentiated in a number of sub sectors, each of them having a particular mission (Dusdal et al., 2020). In what follows, we focus on the public sector only and consider the development of the institutional landscape until 2016 to make sure that each institution existed at least for two years within our observation period.

Universities (UNI) In terms of publication output, universities are the largest sector in the German research system. In 2016, the sector consisted of 96 universities. The number of scientific staff at all universities (including universities of education and theological colleges) was 286,691 full time equivalents in 2018 (Statistisches Bundesamt, 2019) and the budget of all universities (including medical and health science institutions at universities) summed up to 48.989 billion Euro (Statistisches Bundesamt, 2020b). Given that universities follow at least the two missions research and teaching, only parts of the budget were spent on research. Governing bodies are the federal states that contribute the majority of 75% of the funding of the universities. The federal government is involved in the funding of universities via different programmes. With a few exceptions of smaller universities, the research portfolio of the large majority of the universities cover many disciplines and subjects often ranging from the natural sciences, life sciences and engineering, to the social sciences and the humanities (Dusdal et al., 2020). Some tensions can also arise from the structural preconditions associated with university governance. On the one hand, German universities are public institutions that are predominantly funded by one of the states. On the other hand, the German constitution guarantees individual members of the universities 'freedom of teaching and research'. These conditions result in low institutional autonomy and high autonomy of the individuals, especially professors (Schimank, 2005). Regarding the advancement of OA, universities tend to be responsive towards specific targets especially when set by the state, while the ability to enforce compliance of the universities' members is low.

Helmholtz Association (HGF) The Helmholtz Association is an umbrella organisation that consists of 21 Helmholtz Research Centres conducting large-scale research. Given that the organisation provides large-scale research facilities and instrumentation that is open to

the use by the international scientific community, it has strong international collaborations. Today's mission of the Helmholtz Association is to contribute to solutions to grand challenges in the fields of 'energy', 'earth and environment', 'health', 'aeronautics, space and transport', 'matter', and 'future technologies' (Goebelbecker, 2005). Therefore, each centre has a disciplinary profile with a strong publication output in specific fields, for example in engineering or health. The volume of public expenditures was 4.404 billion Euro in 2018, of which 90% came from the federal government and 10% from the state in which the research centre is located (Statistisches Bundesamt, 2020a). The staff in research and development at the centres sum up to 32,853 full time equivalents (Statistisches Bundesamt, 2020a). Compared with other German research organizations, the Helmholtz Association and each individual centre tend to have a strong organizational hierarchy. This is also reflected in the OA activities of the Helmholtz Association which are coordinated by a central unit—the Helmholtz Open Science Office.[8] The tasks of the office focus on policy and support, however, OA facilities such as institutional repositories or publication funds are managed by each research centre.

Fraunhofer Society (FhS) The mission of the Fraunhofer Society is to perform application- and technology-oriented research for the industry but also for the service sector and the government (Mitchell, 1998). It aims to bridge the innovation gap of basic research and supports a rapid commercialisation of technology. Established in 1949, it is organised in a number of Fraunhofer institutes, each of these being centres of excellence in a well-defined area of research. In view of its application-oriented mission, it is not surprising that the main outcomes of the Fraunhofer institutes do not necessarily address the scientific community in the format of scientific publications but also consists of patents as a means to transfer knowledge. The autonomy of the institutes is high within a framework of uniform rules and contracts. The Fraunhofer Society is a non-profit organisation that rests on three pillars of funding: Institutional funding (roughly 30%), contract research and publicly funded research projects (roughly 70% together). In 2018, the volume of public expenditures for the Fraunhofer Society was 2.562 billion Euro and the number of staff in research and development (full time equivalents) was 18,206 (Statistisches Bundesamt, 2020a). This study covers all 68 Fraunhofer institutes that existed in Germany in 2016 but also other facilities of the Fraunhofer Society such as Fraunhofer Working Groups, Fraunhofer Alliances and Fraunhofer Centres. The Fraunhofer Society has a uniform OA policy[9] and provides a central repository that is open to all members of the society.

Max Planck Society (MPS) The Max Planck Society is an independent non-profit organisation with the mission to support research excellence in fundamental research. Each of the Max Planck Institutes (MPIs) should be organised and run according to the Harnack principle, that can be understood as the guiding idea to build institutes around outstanding researchers. They are selected by the council of the Max Planck Society and make all decisions hiring staff (Peacock, 2016). Today, building on the original idea, MPIs are run by 'collegial directorships' involving two to five directors. Because of the Harnack principle, each institute has a comparatively narrow subject focus and the publication output may therefore be represented by the specific publication culture of a subject field. Moreover, the autonomy of the Max Planck Society and each institute are both high. Roughly 90% of the budget of Max Planck Society are public funds that summed up to 1.993 billion € in 2018,

[8] https://os.helmholtz.de/.

[9] https://www.openaccess.fraunhofer.de/en/open-access-strategy.html.

of which 50% comes from the federal government and 50% from the federal states. The society currently employs 15,736 full time equivalent staff in research and development (Statistisches Bundesamt, 2020a). With the Max Planck Digital Library (MPDL), the Max Planck Society has a central unit that is responsible for the provision of scientific information to all institutes of the society. It is an OA proponent and supplies infrastructures and services including repositories and publication funds for all members of the society. This study covers all of the 86 MPIs located in Germany in 2016 but also other organisational entities like Max Planck centres, networks and groups.

Leibniz Association (WGL) The Leibniz Association is an incorporated society of independent institutes and research organisations that derive half of their basic budget from the federal government while the other half comes from the federal state in which the institute is located. Due to the historical development, the Leibniz Association consists of a large diversity of institutes and organizations with different missions. It includes institutes that are dedicated to basic and applied research but also organisations with the purpose to maintain research infrastructures (like museums, libraries and collections) and to provide research-based services. A precondition of an organisation to be incorporated into the Leibniz Association is excellence in performance regarding their mission and interest and relevance of the work for the federal states as a whole (Wissenschaftsrat, 2013). Most of the institutes have a relatively narrow focus regarding the topics studied and—in many cases—a clear orientation towards a discipline. In 2018, the organisations of Leibniz Association received an overall volume of 1.807 billion Euro of public funds and the number of staff in research and development (full time equivalents) was 12,946 (Statistisches Bundesamt, 2020a). OA as a field of action is put forward by a mixture of central and decentral approaches. On the level of the association, the Leibniz Association has uniform OA guidelines and policies, an aggregator for publications archived in repositories of the institutes (LeibnizOpen[10]) and a central publication fund. On the level of the institutes, institutional repositories are provided. This study covers all 95 entities of the Leibniz Association with a research mission that were part of the association in 2016.

Government Research Agencies (GRA) of the federal states: The mission of the Government Research Agencies is threefold. First, the aim of the institutions is to conduct research, second, they provide policy advice, and third, they are involved in state regulation, standardisation, and marketing authorisation (Barlösius, 2010). The category Government Research Agencies was created by the ministries of the federal states and in most cases Government Research Agencies are subordinate agencies of a ministry. The authoritative list of all Government Research Agencies is included in the Bundesberichte Forschung und Innovation, published by the Federal Ministry of Education and Research (BMBF, 2016). In consideration of their mission, research in Government Research Agencies is oriented towards the demand of the governing ministry and is therefore problem-oriented, applied, and in many cases also interdisciplinary. The profile of each agency focuses on a specific topic, for example, traffic and transportation, materials research, labour market research, or nutrition. In 2018, the overall budget of the Government Research Agencies was 2.370 billion Euro including 1.196 billion Euro for research and development and the number of staff in research and development (full time equivalents) was 9747 (Statistisches Bundesamt, 2020a). Regarding OA, the "AG Ressortforschungseinrichtungen", a network of many Government Research Agencies, mentions the goal of OA to publications in one of their statements (AG Ressortforschungseinrichtungen, 2013) but the support of OA takes

[10] http://www.leibnizopen.de/.

place on the level of individual agencies. This study covers all 67 Government Research Agencies with a research mission that were mentioned in BMBF (2016).

Open access in Germany

This empirical study focuses on OA journal articles from authors affiliated with German universities and non-university research institutions between 2010 and 2018. During that period, support structures for OA publishing in Germany broadened. Similar to the international situation (Pinfield, 2015), research policies and measures targeted OA through journals and repositories simultaneously. To contextualize our investigation, we will briefly review major OA advancements in Germany.

Policy context

German universities, research organisations and funders were among the first to officially support the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities which was initiated by the Max Planck Society in October 2003. Shortly after, they began to outline their strategies (Schmidt & Ilg-Hartbecke, 2009). The German Research Foundation (Deutsche Forschungsgemeinschaft—DFG), the largest German research funder, made OA an integral part of its funding policies. But also many universities and research organisations have committed to OA since then. To coordinate OA policies and activities, the Alliance of Science Organisations in Germany[11] formed the priority initiative "Digitale Information"[12] in 2008. Although not directly involved in this initiative, many states and the federal government committed to OA as well. As one of the first federal states in Germany, Berlin launched its “Open Access Initiative” in 2015, providing structural support to implement its policy including annual OA monitoring (Voigt et al., 2018). At the federal level, the German copyright law introduced a clause in 2014, which allows authors to make their final accepted manuscripts freely available through, for instance, an OA repository, if the results originated from mainly publicly-funded research activities, and if the work appeared in a periodical.[13] The embargo period is twelve months after publication regardless of the publisher’s policy. In 2017, the Federal Ministry of Education and Research announced funding for a country-wide OA monitoring effort based at the Forschungszentrum Jülich.[14]

Journal-provided OA (Gold OA)

In Germany, universities and research organisations have promoted OA via both journals and repositories. The DFG has strongly influenced how publication in OA journals is supported. Since 2007, the funder has provided financial support to OA journals affiliated

[11] Members of the Alliance are the Alexander von Humboldt Foundation, the German Academic Exchange Service (DAAD), the German Research Foundation (DFG), the Fraunhofer-Gesellschaft, the Helmholtz Association, the German Rectors’ Conference (HRK), the Leibniz Association, the Max Planck Society, the German National Academy of Sciences Leopoldina and the German Council of Science and Humanities (Wissenschaftsrat).

[12] https://www.allianzinitiative.de/.

[13] https://www.bildung-forschung.digital/de/zweitveroeffentlichungsrecht-2667.html.

[14] https://open-access-monitor.de/.

with a German university or society, targeting both new-born and established journals that intended to transfer to an OA business model (Fournier, 2007). In 2011, the first DFG-funded university-wide publication funds to cover publication fees, often called article-processing charges (APCs), began to operate. Before this, only a few institutional funds existed (Eppelin et al., 2012). Combined with financial support to pay APCs, the DFG enforced a set of criteria that resulted in similar policies regarding the reimbursement of publication fees across German universities (Fournier & Weihberg, 2013). According to these criteria, publication fee spending was capped at € 2000 per article and hybrid journals were excluded from funding.

Although non-university research organisations had not been able to apply for DFG support to cover publication fees, they aligned their efforts (Bruch et al., 2015). For example, the Forschungszentrum Jülich and the Helmholtz-Zentrum Dresden-Rossendorf, both affiliated with the Helmholtz Association, excluded articles in hybrid journals from funding according to the Open Access Directory.[15] Likewise, the Leibniz Association set up a dedicated OA fund supporting articles published in fully OA journals. In terms of workflows, the Max Planck Society acted as a role model by handling publication fee spending centrally through the Max Planck Digital Library (MPDL) (Schimmer et al., 2013; Sikora & Geschuhn, 2015). Because of the aligned funding criteria for publication fees, Germany's spending profile differed from that of the United Kingdom or Austria where no price caps were in place, and where the hybrid model was funded extensively (Jahn & Tullney, 2016; Pinfield et al., 2016).

German funders, universities and research organisations generally discouraged hybrid open access, but some have piloted this route. Between 2007 and 2012, the University of Goettingen had an agreement for hybrid open access in Springer journals (Mueller-Langer & Watt, 2018). The Max Planck Society participated in the Springer Compact program, a read and publish agreement (Olsson et al., 2020). Launched in 2017, the German BMBF Post Grant Fund did not explicitly exclude hybrid open access. Besides, a survey suggests that German authors were able to make use of discretionary funds to finance open access publications in hybrid journals (Van der Graaf, 2017).

German funders, universities and research organisations increasingly negotiate OA agreements with major publishers. Germany has a long tradition of joint licensing of digital collections, both from a national and a federal state perspective. These agreements have evolved over time, starting from subscription and archiving licenses, and increasingly take the emerging OA models into account. Since launching of the priority initiative "Digital Information" in 2010, national licensing of electronic journals was further developed into opt-in consortia-based "alliance licenses" that allowed participating institutions to deposit articles from their institutions immediately or after an embargo period. From 2014 onwards, the Max Planck Society, the Helmholtz Society and many universities have contributed financially to the international SCOAP^3 consortium,[16] which aims at converting subscription-based high-energy physics journals to OA (Kohls & Mele, 2018). Relating its activities to the international OA2020 initiative, which calls for a transparent approach to re-allocate budgets currently spent for subscriptions to OA business models, the DFG introduced the funding program "Open Access Transition Agreements" in 2017. It aims at library consortia negotiating Germany-wide transformative agreements with publishers. Likewise, the German DEAL consortium, representing more than 700 universities and

[15] http://oad.simmons.edu/oadwiki/OA_publication_funds.

[16] https://scoap3.org/.

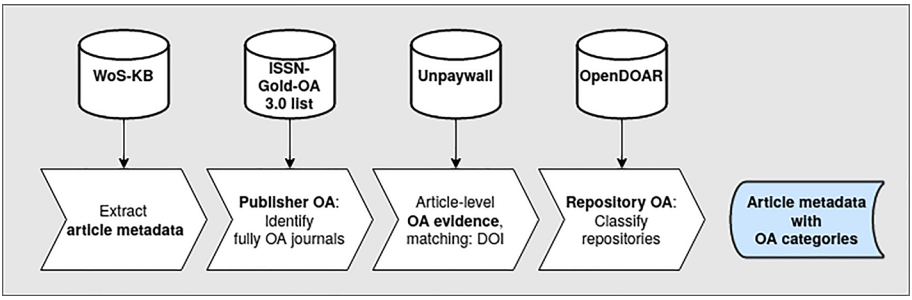

Fig. 1 Study design: schematic display of gathering, matching, and preprocessing of data (read from left to right). First, article-level data was obtained from the Web of Science in-house database of the German Competence Centre for Bibliometrics (WoS-KB) including its standardised affiliation information. Next, fully OA journals were identified using the ISSN Gold-OA 3.0 list. Article-level OA evidence from Unpaywall (data snapshot from February 2020) was added by DOI matching. After that, Unpaywall's OA evidence was complemented with data from OpenDOAR to further differentiate OA categories

non-university research institutions, planned to transact transformative agreements with the leading publishers Elsevier, Springer Nature and Wiley. So far, Germany-wide transformative agreements were successfully negotiated with Springer Nature, Wiley, IOP Publishing and Cambridge University Press, while negotiations with Elsevier stalled (Vogel, 2019). Presumably, these Germany-wide agreements, which all came into effect after 2018 (2018 is the end date of our investigation) will lead to an increased proportion of OA journal articles published by corresponding authors affiliated with German universities and research organisations.

Repository-provided OA (Green OA)

Complementary to the journal route, OA via repositories, interoperable online archives for scholarly works, has been endorsed by OA policies. According to Schmidt and Ilg-Hartbecke (2009) already half of research-intensive universities in Germany maintained a repository before 2010. Today, most universities and research organisations provide an institutional repository; at the same time, they encourage self-archiving in subject-specific repositories. Again, the DFG provided support for the launch and networking of repositories. In Germany there are only very few institutional OA policies which mandate deposit of publications in repositories. The University of Konstanz, for instance, has required authors affiliated with the university to take advantage of the German copyright reform to self-archive accepted manuscripts, leading to a yet unresolved legal dispute between the University of Konstanz and university lecturers.

The German repository landscape is characterised by a high level of standardisation. Following the international Open Archive Initiative (OAI) (Lagoze & Van de Sompel, 2003), the German Initiative for Networked Information (Deutsche Initiative für Netzwerkinformation—DINI) has promoted web standards to ensure that OA literature in repositories is discoverable, preserved and exchangeable. Since 2004, DINI has certified repositories against a comprehensive set of criteria (Müller & Schirmbacher, 2007). From the beginning, these criteria have been aligned with the OAI standards and related European standardisation efforts driven by the EU-funded projects DRIVER (Lossau & Peters, 2008), OpenAIRE (Schirrwagen et al., 2013) and the Confederation of Open Access

Repositories (COAR e.V.). The German Bielefeld Academic Search Engine (BASE)[17] is a prominent example demonstrating the outstanding importance of the OAI standards for OA adoption in Germany and worldwide (Pieper & Summann, 2006).

Methodology

Data were assembled from multiple sources to obtain the OA profile of German universities and non-university research organisations between 2010 and 2018 as described in Fig. 1. Table 1 presents the study's OA classification and operationalisation methodology.

The Web of Science in-house database maintained by the German Competence Center for Bibliometrics (WoS-KB), 2019 version, was used to determine the publication output of German universities and non-university research institutions. The main advantage of the WoS-KB for the purpose of our study is its disambiguated address information (Donner et al., 2020; Rimmert et al., 2017), which not only allowed obtaining the publication output at the institutional level, but also to interlink them to the specific sectors of the German research system. The institutional disambiguation authority file was developed at the Institute for Interdisciplinary Studies of Science at Bielefeld University and works as a central component with a "near-complete national-scale coverage" of Germany's institutions represented in the Web of Science (Donner et al., 2020). Accordingly, Donner et al. (2020) reported a very high accuracy. The disambiguated affiliation system for German institutions can, thus, serve as a gold standard for institution name disambiguation. Technically, the address information was generated via a rule-based affiliation disambiguation system. For our study, we used address information for about 2000 German academic institutions, distributed across all research sectors.

Following Peter Suber's seminal work (2012), we distinguished between OA provided by journals ("Gold OA") and repositories ("Green OA"). These top-level categories were further differentiated into OA subtypes by a combination of different data sources:

- **Fully OA Journal**: To identify articles in fully OA journals the ISSN-GOLD-OA 3.0 list (Bruns et al., 2019) and Unpaywall's journal classification (Piwowar et al., 2019) were combined to create an exhaustive list. We were able to identify 1986 fully OA journals that included at least one article authored by a researcher at a German research institution, according to WoS-KB. Of these, 1322 were classified as such by both data sources, 158 only by Unpaywall, and 506 exclusively by the ISSN-GOLD-OA 3.0 list.
- **Other OA Journal**: Based on article-level evidence from Unpaywall, articles were assigned to the category 'Other OA Journal' if Unpaywall's field 'host type' (specifying if a resource was found on a publisher hosted platform or in a repository) was tagged as 'publisher' but the journal was not fully OA. We decided not to apply Unpaywall's classification of "hybrid" and "bronze", because of ongoing conceptual and methodological controversies about how to determine hybrid journals and immediate open access articles published in them (Akbaritabar & Stahlschmidt, 2019; Piwowar et al., 2019). They led to ongoing changes to Unpaywall's methodology relative to hybrid open access, making it difficult to compare uptake rates across studies based on

[17] https://www.base-search.net/.

Table 1 Study design: OA classification

Host	Variable	Definition	Evidence source
Journal-provided OA (Gold OA)	`full_oa_journal`	Article published in a **fully OA journal**	`(host_type="publisher"` AND `journal_is_oa=TRUE) OR` journal listed in ISSN-GOLD-OA 3.0
	`other_oa_journal`	Article full text provided by a **non-fully OA journal**	`host_type="publisher"` AND NOT (`journal_is_oa=TRUE OR` journal listed in ISSN-GOLD-OA 3.0)
Repository-provided OA (Green OA)	`opendoar_inst`	Article full text provided by an **institutional repository** registered in OpenDOAR	Article full text domains from Unpaywall Data with `host_type="repository"` OpenDOAR repository domain
	`opendoar_subject`	Article full-text provided by a **subject-specific repository** registered in OpenDOAR	
	`opendoar_other`	Article full-text provided by **other types of repositories** registered in OpenDOAR	
	`other_repo`	Article full-text provided by a **repository**, which is **not registered in OpenDOAR**	`host_type="repository"` AND full text domain not listed in OpenDOAR

For both of the main OA routes (Gold and Green OA), the categorisation into subtypes, their definition as applied in this paper, and the evidence bases used for the distinctions are displayed

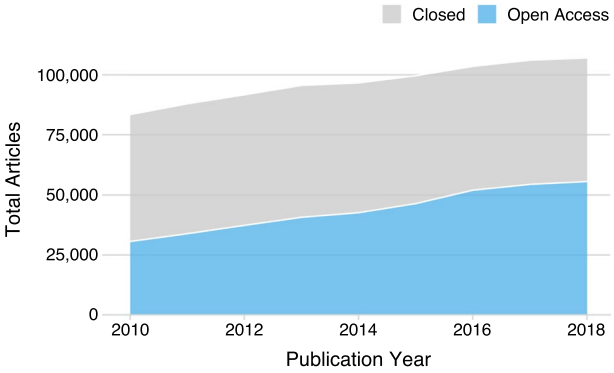

Fig. 2 Open access to journal articles from German research institutions by year (see *Table S.1.1* in the supplementary material for the explicit numbers)

these categories. After the time of writing, for instance, Unpaywall revised its approach to distinguish between "hybrid" and "bronze" in October 2020.[18]

- **Repository-provided OA**: To identify articles in repositories we, again, built on Unpaywall. Domains from repository full-texts links were extracted and matched with the Directory of Open Access Repositories (OpenDOAR), a comprehensive registry of repositories supporting the OAI standard. Using the OpenDOAR repository classification, we distinguished between **institutional**, **discipline-based**, and **other types** of repositories. If a domain was **not listed in OpenDOAR**, repository full-texts were classified as "other".

The analysis was carried out in March 2020 using the most current datasets available at that time. We selected the document types "articles" and "reviews" from the Web of Science database editions Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), and Arts and Humanities Citation Index (A&HCI). We applied full counting of all authors, meaning that articles affiliated with authors from multiple institutions were counted once for each associated institution. In total, 871,922 articles from WoS-KB met our selection criteria. Of those, 95% had DOIs and 94% were matched to articles in Unpaywall. 5966 DOIs in our WoS-KB sample could not be matched. Automated checking with the Crossref API using rcrossref (Chamberlain et al., 2020) revealed that 57% of non-matched DOIs did not resolve, while 43% were not registered by Crossref, but other agencies like DataCite. Unpaywall only tracks Crossref DOIs.

For analysis, we extensively used the tidyverse R package family (Wickham et al., 2019). Source code used for data gathering, analysis and validation including notebooks are available via GitHub.[19]

[18] https://support.unpaywall.org/support/solutions/articles/44001867302-unpaywall-change-notes.
[19] https://github.com/subugoe/oauni.

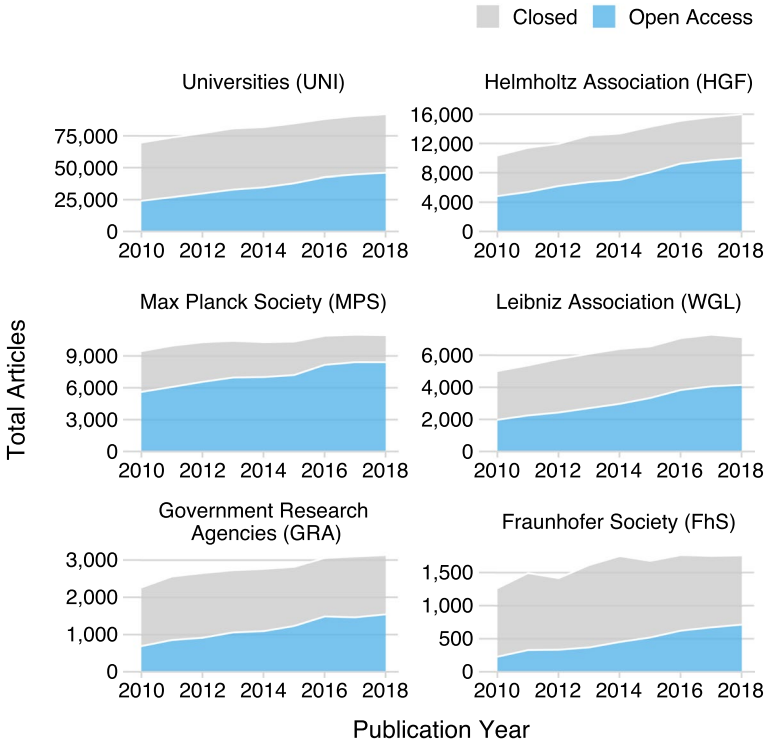

Fig. 3 Development of the number of OA/closed access articles, by sectors (2010–2018). (The exact numbers for each sector and year can be found in *Table S.1.2* in the supplementary material.) Note that scales for the vertical axes differ, since the total publication output varies significantly among sectors

Results and discussion

OA fraction of the German publication output

In a first step, we analysed how the overall OA share of the German research system developed over the time period from 2010 until 2018. The following Fig. 2 displays the number of publications with addresses of German research institutions and highlights the freely accessible subset. The overall OA share was 45% considering all years collectively. This finding is in line with results from Robinson-Garcia et al. (2020), who reported 43% as the global median OA share of publications from universities in the period 2014–2017, with a slightly higher share for German universities. Piwowar et al. (2018) reported a slightly lower OA percentage of 36% for a sample of 100,000 articles registered within the Web of Science that were published between 2009 and 2015.

As Fig. 2 shows, the total number of articles, as well as the number of OA articles increased constantly over time. The absolute number of toll-access articles was quite stable with a slow increase from 52,803 in 2010 to 54,873 in 2013, and decreasing again from that point onwards to 51,430 publications in 2018. Since the number of OA articles increased continuously from 30,664 publications in 2010 to 55,649 in 2018, the relative proportion of OA articles rose from 37% in 2010 to 52% in 2018.

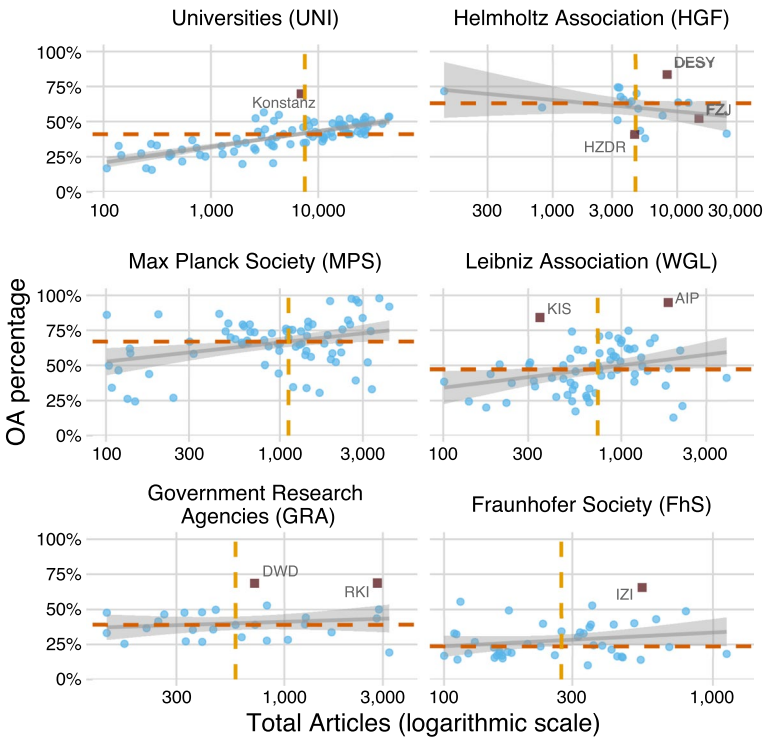

Fig. 4 OA shares and publication output of German research institutions with at least 100 publications in 2010–2018, grouped by sectors. Solid grey lines are obtained by linear regression within the sector, shaded grey areas are pointwise symmetric 95% t-distribution confidence bands. Dashed lines represent the median values of the OA share (red) and the publication output (orange) of the sectors. Labelled squares in darker colour highlight institutions mentioned in the text. Scales of the x-axes vary across subplots in order to adapt to the different publication volumes

As an answer to research question RQ1, we were able to establish that the OA fraction of the publication output of German universities and non-university research institutions has been rising continuously over the observed time period from 2010 to 2018, confirming the international trend.

Differences between research sectors

In a next step, the development of the OA shares are analysed for the different sectors (universities, non-university research institutes like MPS or WGL institutes, and government research agencies) of the German research system separately. The results are displayed in Fig. 3.

Two results of the cross-sector comparison are highlighted: First, the total publication output varied strongly between sectors. The differences in the publication output do not result from the different sizes of the sectors (in terms of budget and staff) only but also reflect the different missions of the sectors. The publication outputs of the sectors oriented towards basic research (like UNI, MPS, and HGF) were considerably larger than those of sectors with a practise-oriented mission like GRA and the FhS. Second, a similar trend can

be found with respect to the OA shares across all the sectors. Again, sectors with an academic orientation and basic research focused mission outperformed the two more practice-oriented sectors regarding the adoption of OA. Of all sectors, the MPS had the highest OA share over the whole period, rising from 59% in 2010 to 77% in 2018. The HGF shows a strong rise both in the overall publication output (from 10,365 publications in 2010 to 15,996 publications in 2018) and in the OA share that rose from about 47% in 2010 to about 63% in 2018. The example is of particular interest as it shows that an increase of the publication output does not necessarily have to happen at expense of the OA share. Compared with these numbers, the fraction of OA publications of the two sectors with practise-oriented missions were low (41% for GRA and 29% for FhS).

In order to deepen the understanding of OA within the German research landscape, the OA shares of individual institutions, grouped by sector were calculated. The analysis was restricted to institutions with a publication output of at least 100 publications in the period 2010–2018 and excluded administrative facilities as well as residual and aggregating categories. Of the 444 institutions in total, 320 meet these conditions, while 124 institutions with a cumulated volume of 6259 articles were excluded from this step of the analysis.[20]

Figure 4 displays the results.

A comparison of the scatter plots of the different sectors suggests that the distributions are not determined by a single factor but by a combination of different factors.

For UNI, the spread around the linear trend line was very low, indicating that the OA shares were partly determined by its size, as measured by their overall publication output—universities with larger publication outputs tended to have larger OA shares. Outliers with above-average OA shares were universities that strongly support OA or that are known as OA pioneers in Germany. An example is the University of Konstanz with the highest overall OA share of 70% among all German universities. Compared with the other two basic research-oriented sectors (MPS and HGF) the OA share of UNI was comparatively low. Possible reasons might be, on the one hand, that researchers based at universities enjoy a high degree of autonomy guaranteed by the German constitution that makes it difficult for the management to enforce compliance with OA policies. On the other hand, research at German universities covers a large variety of disciplines and fields, including those with both high and low adoption of OA.

Evidence for the influence of disciplinary publication cultures on OA shares can be drawn from the scatter plot of another sector of the research system, the MPS. Following the divisions of the four quadrants separated by the two median lines, physics and astronomy institutes were located in the upper right corner with a high publication output and a high OA percentage. Researchers in this field traditionally tend to publish preprints on subject-specific repositories and the landscape of the journals are characterised by a high level of openness (Taubert, 2019). In the upper left quadrant with similarly high OA shares but with lower publication counts, institutions with a life science profile dominated. In the lower left quadrant humanities' and social sciences' institutions accumulated, having had a lower publication output in journals covered by WoS-KB and a lower OA share. Lastly, the lower right quadrant, characterised by an above-average number of publications, but an OA share lower than median, was occupied mostly by institutions with a focus in materials research.

[20] For the aggregated data of all individual institutes and categories, see *Tables S.2.2—S.2.7* in the supplementary material.

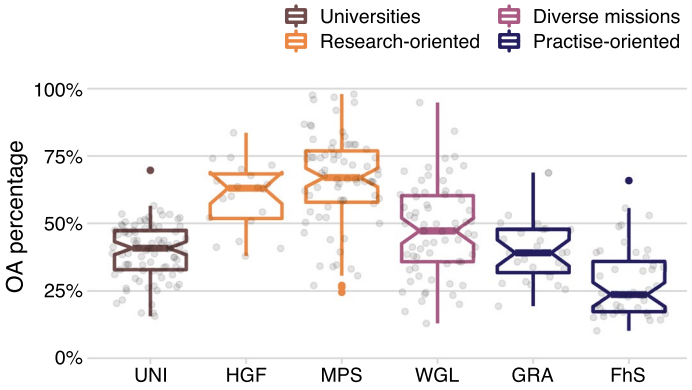

Fig. 5 OA shares of German research institutions with at least 100 publications in 2010–2018, grouped by sectors. The colour of the boxes groups sectors into universities with a typically high total journal publication output and diverse subject profile, research-oriented institutes with a medium journal publication output and often a specific disciplinary focus, practise oriented institutions with a comparatively low journal publication output, as well as sectors with diverse missions of their institutions. Points display the OA shares for individual institutions. Bars show the median, boundaries of the boxes are at first and third quartiles. Whiskers extend to the furthest value no further than 1.5 * IQR from the hinge, where IQR is the interquartile range, or distance between the first and third quartiles. Outliers are displayed separately as coloured points. Notches indicate approximate 95% confidence intervals for the median values. Non-overlapping notches imply a strong indication that median values differ significantly

In the case of the HGF the distribution also seems to be influenced by the disciplinary publication culture. The majority of institutions with an OA percentage above the median value were located in the natural and life sciences. The highest OA share (84%) of all Helmholtz institutes was registered for the Deutsches Elektronen-Synchrotron (DESY) a large-scale research facility in (particle) physics. The plot showing the institutions of the HGF also suggests that disciplinary publication cultures have had a stronger influence on the OA share than institutional support. For example, the Jülich Research Centre (FZJ) and the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) both support publication in fully OA journals with their publication funds and provide repositories for self-archiving but their overall OA percentages were below the median value for this sector (52% and 41% compared to a median value of 63%).

The FhS has other major output formats aside from journal publications, like patents and technology products and reports. However, those are not covered by this analysis based on the Web of Science. The comparable low OA share of most of the Fraunhofer institutes may reflect the more application-oriented specific mission of the sector.

An interpretation of the results for the WGL and the GRA is less straightforward, since these two sectors comprise heterogeneous institutions regarding their missions and orientations. However, the disciplinary publication culture again seems to play a certain role also here. The two leading institutions in each sector, namely the Leibniz-Institute for Astrophysics Potsdam (AIP), the Leibniz Institute for Solar Physics (KIS), the Robert Koch Institute (RKI), and the Deutscher Wetterdienst (DWD) can all be attributed to the natural sciences, and physics in particular, as well as to the life sciences.

Figure 5 quantifies the observations regarding the variability of OA shares within sectors that we already made in Fig. 4. Using the non-overlapping of boxplot notches as an approximate measure of significant differences in median, we deduce that the two research-oriented sectors, HGF and MPS, had significantly higher median values for the

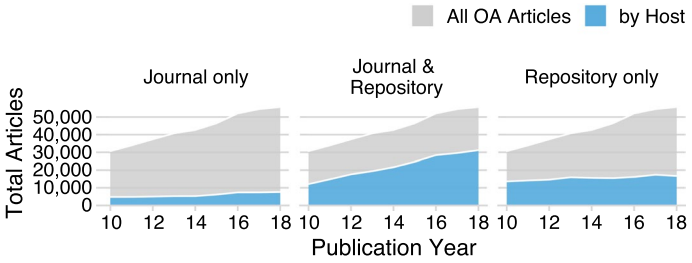

Fig. 6 Development of the number of articles per OA-type and their overlap. Highlighted in blue are the number of articles per OA host type ('by Host') with articles made available only via a journal on the left, articles available only in repositories on the right and the overlap, that is, articles openly accessible via both a journal and a repository, in the middle. Grey area shows the remaining OA articles. Exact numbers can be found in *Table S.1.3* in the supplementary material

OA percentage than the other sectors. On the other end of the spectrum, the more practise-oriented institutes of the FhS had a much lower OA percentage than all other sectors. UNI with a typically very diverse disciplinary profile, and WGL and GRA with their diverse primary missions all had intermediate levels of their median OA percentage. Furthermore, we can confirm the observation that the variation of OA percentages within the sector of UNI is very low, whereas for the WGL the diverse strategic focuses might be a key factor explaining the high spread of OA shares.

Regarding research question RQ2, we found large differences in the degrees of OA adoption for the research sectors of the German research landscape. These differences may originate from the diverse disciplinary profiles of the research institutions as well as differing key missions. Moreover, the different orientations toward basic research versus application in practise or supply of infrastructure typically amount to vastly different importance of journal publication as a research output. However, more rigorous investigations are necessary to determine the influence of the different factors.

Prevalence of OA categories

As outlined previously, there are several ways of providing OA to publications. In this section, research question RQ3 is addressed as the prevalence of the most widespread OA routes is investigated: OA via repositories (Green OA) and via journals (Gold OA). In the case of Gold OA, we further distinguished between articles in fully OA journals and other types of OA provided by journals (e.g., delayed, hybrid and promotional OA). In the case of repositories, we distinguish between disciplinary, institutional, and other OpenDOAR-listed repositories as well as sources not registered within OpenDOAR. The OA categories are non-exclusive, that is, an article might be counted for several categories. Articles were fully counted in every category they appear in. Hence, numbers do not sum up to the total number of articles considered in this study, and percentages do not sum up to one hundred per cent.

As a first step, the relevance of the two main OA types is analysed in Fig. 6.

The most striking observation is that the majority of openly accessible journal articles (51% of all OA articles over the whole observation period) were available through both types: via the journal and also via at least one repository. Moreover, this overlap also shows the strongest increase over time, from 12,136 articles in 2010 to 31,237 in 2018. Articles

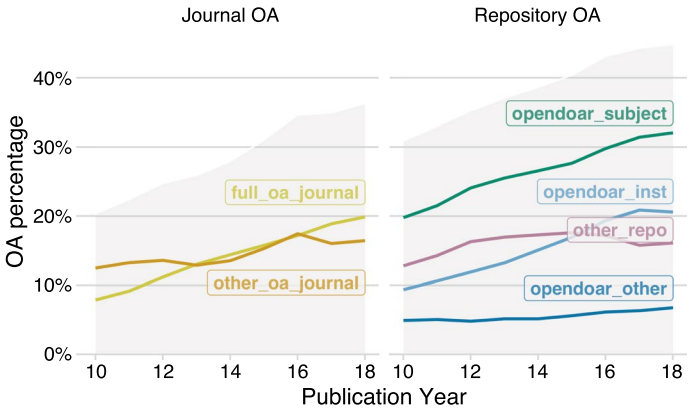

Fig. 7 Development of the percentage of journal articles per OA category. Categories are non-exclusive, that is articles may be counted for more than one category. Grey area displays the total percentage of the major OA type (journal or repository). Exact numbers can be found in *Table S.2.1* in the supplementary material

that were available exclusively via a journal are the minority, yet the numbers have risen strongly over time from 4860 articles in 2010 to 7668 articles in 2018. In addition, there is a relatively steady amount of around 15,500 articles published every year which was OA exclusively via a repository.

A closer inspection of the data reveals that of the articles which were OA exclusively via a journal (highlighted in blue as 'by Host' in the left column in Fig. 6), only 33% were published in fully OA journals, while the remaining 67% were other journal provided OA types like delayed, hybrid and promotional OA. This distribution strongly differs from the second group, where OA was provided via journals and repositories (highlighted in blue as 'by Host' in the middle column in Fig. 6). Here, more than half of the articles (54%) were published in fully OA journals. In other words: it is more likely for an article in a fully OA journal to be archived on a repository than for an article where journal-provided OA follows a different model. Robinson-Garcia et al. (2020) suggest that this partially might be a result of indexing in PubMed Central including Europe PMC.

Turning to the repository categories, and keeping in mind that articles may be deposited in more than one repository, in both cases (overlap and exclusively repositories), subject-specific repositories contributed the largest share. However, while little more than half (54%) of the articles that were OA exclusively via a repository (highlighted in blue in the right column of Fig. 6) were deposited on a subject-specific repository, this was the case for almost 80% of articles in the overlapping group. A similar observation can be made for the residual category 'other_repo' with 30% occurrence in the exclusive repository group, and 49% in the overlapping group. Institutional repositories (around 40%) as well as other OpenDOAR registered repositories (around 14%) appeared equally often in both groups.[21]

Figure 7 shows that of all OA sub-categories, journal- and repository-provided, subject-specific repositories as classified by OpenDOAR were the most prevalent OA subtype in each year of the period analysed in this study. This is in contrast to findings

[21] For details see *Table S.3.3* in the supplementary material.

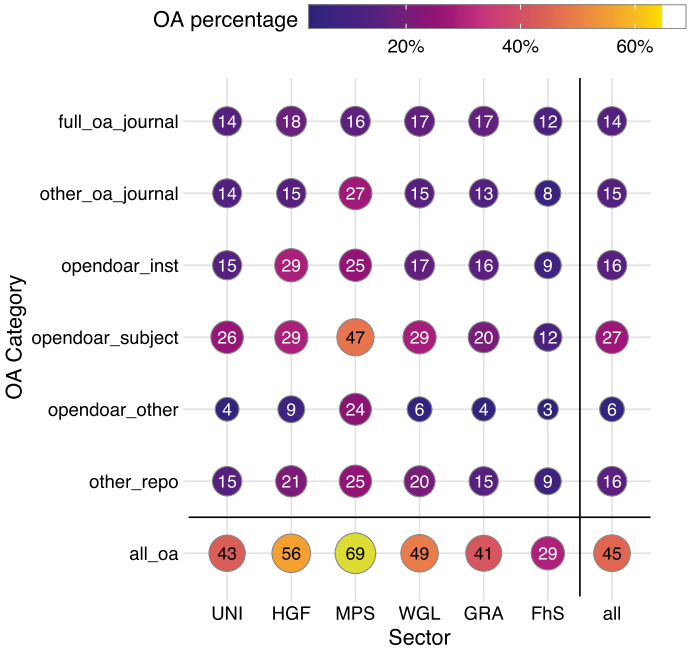

Fig. 8 OA shares per category and sector of articles published between 2010 and 2018. Colouring and size of the points displays the percentage in the respective category. Grey numbers display the percentage value. Note that categories are not exclusive, so percentages do not necessarily add up to 100. The underlying numbers can be found in *Table S.2.1* in the supplementary material

from earlier studies that base their analyses on the field best OA location of Unpaywall (Martín-Martín et al., 2018; Piwowar et al., 2018; Voigt et al., 2018).

Regarding the different journal OA subtypes, three findings are highlighted here: First, there was a growth of the percentage for both articles in fully OA journals and for other OA types provided by journals ('other_oa_journal') in the observation period. Second, the growth of the percentage of articles in fully OA journals was larger and at first glance it seems that this sub-category has become more important than other OA types provided by journals. However, these trends should be interpreted carefully as there was a notable drop in the percentage of other OA types provided by journals in the years 2016–2018. This is most likely caused by delayed OA journals where some or all articles of a journal are made available after a certain embargo period which can extend up to several years. Articles from these publication years may therefore not have been openly accessible at the time of analysis but will become OA in the near future. Third, most articles in fully OA journals were published with Springer Nature and Public Library of Science (PLOS). However, the strongest increases over time, mirroring the overall increment in this category, were found for Springer Nature, Frontiers Media SA, and MDPI AG. Publication volumes in PLOS grew from 827 articles in 2010 to 3086 in 2013 and from then on continuously decreased, though they remained at a generally high level: in 2018, there were still 1774 articles published by German research institutions in PLOS journals. Presumably, this can be explained by the general trend. Since 2014, a decline in the number of articles published in PLOS ONE was observed, while

the number of articles published by the competing mega journal Scientific Reports has grown since then (Spezi et al., 2017).

Regarding OA provided by repositories and its subtypes, we stress three main findings: First, deposition in subject-specific repositories ('opendoar_subject') was, in terms of OA share, by far the most important subtype. There are no hints that this situation will change in the near future as there has been a sustaining growth of the OA share of this subtype. A more detailed look into the data reveals that the gain for subject specific repositories can be attributed mostly to the arXiv and PubMed Central including Europe PMC. This suggests that a few disciplinary publication cultures impacted the continuing relevance of this OA publication practice. Second, there was a notable drop in the share of articles openly accessible via residual repositories not registered with OpenDOAR ('other_repo') in the years from 2016 to 2018. This decrease in recent years is almost entirely caused by records found on Semantic Scholar, accounting for almost 83% of all articles in this category. The slight decrease in OA publication for institutional repositories ('opendoar_inst') in the last year is presumably caused by delays in deposition due to self-archiving embargoes allowing a deposition only after a certain period. Another reason might be that not all articles were delivered into the institutional repository by the authors themselves immediately after submission or publication. Third, the remaining category opendoar_other shows a continuous increase, which was, however, not as steep as the growth in subject-specific repositories or in fully OA journals.

In the next step, we analysed if the sectors differ regarding the adoption of OA (see research question RQ3). To explore this, OA percentages per category were calculated for each sector. Figure 8 displays the results.

In each sector, the most prevalent type was OA provided by disciplinary repositories. Sectors with a high OA share, like the MPS, had high proportions of OA provided by subject-specific repositories, but also in the case of the FhS that had the lowest overall OA share, this type contributed the most. It is likely that the OA shares of subject-specific repositories reflect to what extent disciplines with strong self-archiving practices contributed to the publication output of the different sectors.

With respect to OA provided by institutional repositories, a comparison of the sectors shows that HGF, an organization with a comparable strong hierarchical structure and a central unit that supports OA, had the highest respective OA shares, while the shares for UNI and FhS were both comparatively low. These findings are compatible with the assumption that the OA share of this type is at least to some extent affected by the relevance of self-archiving in a particular type of organization and the ability of the organization to enforce their members to self-archive their publications. In addition, the secondary publication right granted by German copyright may play a role in the higher share of self-archiving in the non-university sectors as this right applies to mainly third-party funded research only. For articles in the category 'opendoar_other', the particularly high share for MPS is a data artefact caused by an ambiguous classification of the repository of MPS as both "institutional" and "aggregating" within OpenDOAR. Such repositories, which are registered within OpenDOAR but not unambiguously classified, were labelled as 'opendoar_other' in our analysis. The results for the category 'other_repo' are difficult to interpret as this category is dominated by a single repository—Semantic Scholar—that aggregates various content from different sources.

Regarding OA provided by journals, two findings of the cross-sectoral comparison are highlighted: First, the percentage of articles published in fully OA journals seems to be largely independent from the type of organisation as the shares of different sectors do not vary much from the overall percentage of that category for the German research system.

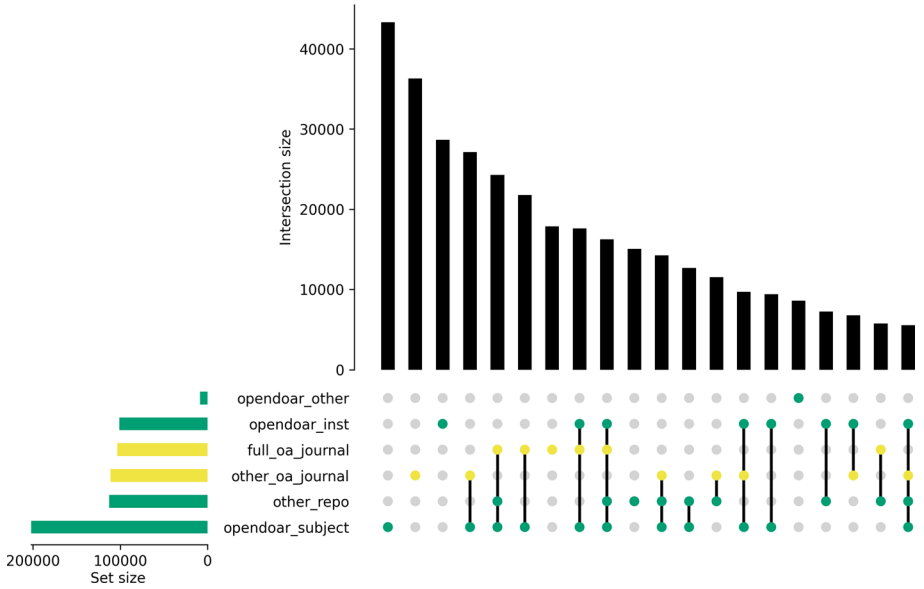

Fig. 9 Overlap of different OA categories (as per schema in Table 1). Only the 20 most prevalent combinations are displayed. Bars on the left show the total number of articles per category. Connected points on the right show combinations of OA categories, represented by coloured circles. Intersecting categories are connected by vertical lines. The upper bar plot displays the number of articles per combination of categories (e.g., the leftmost black bar shows the number of articles for which all locations are classified as subject-specific repositories. The fourth from the left one shows how many articles are openly available in a non-fully OA journal and via a subject-specific repository). Colours correspond to the OA route (via a journal or via a repository)

The results suggest that the shares of the sector may be influenced primarily by the extent to which journals apply a full OA publishing model and not so much by organisational factors.

Second, this finding sharply contrasts to the distribution of the OA shares of the 'other_oa_journal' type, as MPS had a remarkably higher share in this category compared to the overall proportion for all sectors. A more detailed look into the access conditions of the journals that contributed the most to the publication output of MPS in this category shows that the high share to a large extent results from the delayed OA model that is applied by large journals in physics, astronomy, and the life sciences. Therefore, the high OA share of MPS in this category mainly reflects the disciplinary profile of MPS with a strong publication output in these disciplines.

Overlap of OA categories

For 72% of all OA articles in our dataset, Unpaywall tracked more than one OA full text link. In our analysis, we classified each OA location according to our schema in Table 1. As noted, our categories are non-exclusive, i.e. articles that are openly accessible through different means were counted once in each of the categories. In order to quantify this overlap, Fig. 9 displays the most common combinations of OA categories found as upset graph (Lex & Gehlenborg, 2014).

The largest groups were articles available only through a subject-specific repository, followed by articles freely accessible exclusively via a non-fully OA journal and articles on institutional repositories only. Next, several combinations, including articles that were available via a fully or non-fully OA journal as well as through one or more types of repositories, for example on a disciplinary and an institutional repository, followed. These articles were counted fully in each of the OA categories they appeared in. Figure 9 highlights that many articles published in fully OA journals were available through repositories simultaneously, while a larger proportion of OA articles published in otherwise toll-access journals was only available through the publisher website.

With respect to question RQ3, we found that subject-specific repositories are the most prevalent OA type over the whole period on the national level as well as for each sector. However, the percentages for publication in fully OA journals and OA via institutional repositories show similarly steep increases over the observed period. A comparison of the development in different sectors suggests that organisational factors (like centralised or decentralised OA adoption) may influence the share of OA via institutional repositories, and disciplinary profiles may impact the prevalence of OA in subscription-based journals, whereas publication in fully OA journals seems to be affected mainly by the availability of journals offering this publishing model.

Conclusion

Key findings and contributions

Our study presents the first comprehensive empirical study investigating institutional OA uptake in Germany. By reflecting the heterogeneity of German universities and non-university research organisations, this study acknowledges the particularities of the German research landscape. Similar to the international trend and related studies, the overall OA share has grown substantially between 2010 and 2018. However, large variations are observed in terms of productivity, OA uptake and adoption strategies, which can be best explained by the heterogenous research landscape in Germany.

Our study contributes to the evolving body of country-level and institution-specific OA studies. We drew on a quality-assured institutional address coding of the German research landscape based on cleaned and unified Web of Science address information provided by the German Competence Centre for Bibliometrics. Because of this unique affiliation disambiguation effort, we were able to examine not just universities, but also non-university research organisations and their institutions in Germany. Although the evidence-base for OA has evolved in the last years, bibliometric studies on OA still suffer from a lack of standardised methodologies. Most importantly, overlaps between different OA categories need to be addressed. By making these intersections apparent for the German research publication output, our findings demonstrate that prioritising one route over another can lead to misleading interpretations.

Methodological considerations

This study extends existing approaches to address the heterogeneous landscape of OA evidence sources. Combining different journal data sources extends the evidence base for fully OA journals. The inclusion of repository metadata from OpenDOAR further differentiates

Unpaywall's classification of repository-provided OA. Our findings highlight the important role of subject-specific repositories for disseminating journal articles from authors affiliated with German research institutions, followed by institutional repositories. Likewise, our repository classification reflects that standards and interoperability are defining elements of OA repositories. In OpenDOAR, only repositories supporting the OAI protocol are listed, which allows to distinguish whether a full-text is disseminated by a repository complying with this standard or by other means. Most prominently, the recent inclusion of full-text links from the academic search engine Semantic Scholar to Unpaywall as repository-provided OA demands careful consideration when analysing Green OA.

Limitations

This study is not without limitations. Importantly, it must be noted that our focus was on journal articles indexed in the Web of Science only. It is a well-discussed issue in bibliometrics that the Web of Science has a selective coverage and, therefore, likely misses important parts of the scholarly output of an academic institution. But also OA evidence sources are not without limitations. Unpaywall only tracks Crossref DOIs. Therefore, we were only able to obtain article-level OA evidence for Crossref-indexed publications. Other OA discovery solutions like BASE, OpenAIRE and CORE presumably complement Unpaywall's evidence base, in particular, regarding repository-provided content. Although an important part of OA articles was provided through otherwise toll-access journals, we did not further differentiate this OA approach because of the ongoing methodological challenges involved in identifying when an article was made openly available on a journal website. Likewise, it was out of the scope of this study to identify which versions of an article manuscript compared to the peer-reviewed version was deposited within what time frame in a repository.

Outlook

This study is exploratory and time-dependent. Because of the observed large variations in OA publishing patterns between German research sectors, future studies will need to integrate further organisational and subject-specific factors to examine how and to which extent they affect institutional OA adoption. These can be the availability of OA support structures, as well as the disciplinary profile of an institution. Additionally, authorship patterns in terms of author role and collaboration, which were out of the scope of this study, can also contribute to a better understanding of OA adoption at the institutional level.

Recently, Germany and other European countries have started to successfully negotiate transformative agreements with major publishers. Transformative agreements enable corresponding authors to publish OA in subscription-based journals that, in principle, intend to transfer to a full OA business model in future. The journal's belonging to a publisher and corresponding author affiliations therefore become important factors in future bibliometric OA investigations. If these transformative agreements mandate open and standardised scholarly data from the publishers, this will likely extend the evidence-base not just for OA specific, but for all kinds of bibliometric studies.

Overall, our results enable data-driven decision-making in the context of OA in Germany at the level of institutions. Against the background of the ongoing OA adoption in general and the negotiation of transformative agreements in particular, our empirical findings can serve as a baseline to assess the impact of this new publishing model in the future.

Acknowledgements Open Access funding enabled and organized by Projekt DEAL. We are thankful to Neda Abediyarandi, Masoud Davari and Nick Haupka for their assistance with preparing and validating the data and to Nicholas Fraser for helpful comments on the manuscript.

Authors' contributions Conceptualisation and Methodology: AH, NJ, PM and NT; Formal analysis: AH and NJ; Data Curation: AH, NJ and NT; Funding acquisition: NJ, PM, BS and NT; Writing—original draft, review and editing: AH, NJ, PM, BS, NT. Authors are listed in alphabetical order.

Funding This work was supported by the German Federal Ministry of Education and Research within the funding stream "Quantitative research on the science sector", projects OASE (grant number 01PU17005A) and OAUNI (Grant Numbers 01PU17023A and 01PU17023B).

Data availability Article level data is proprietary and can therefore not be published. Together with the submitted version, supplementary material is available on Zenodo via the https://doi.org/10.5281/zenodo.3892951. It consists of three.xlsx files. The file 'S1 Yearly OA shares.xlsx' contains the yearly numbers and proportions of OA articles on the national level and per sector as well as per hosttype (journal or repository provided OA).The file 'S2 Institutional OA shares.xlsx' lists all examined institutions grouped by sector and displays the overall OA percentage, as well as shares of distinct OA approaches for each individual institution. It also contains an overview spreadsheet collating these numbers per research sector. The file 'S3 Data exploration.xlsx' contains additional information on the prevalence of repository domains, unmatched DOIs and distribution of OA categories per hosttype. An interactive dashboard allows visual browsing of key figures and is freely available on https://subugoe.github.io/oauni.

Code availability All codes, scripts and database queries used for data gathering and analysis are openly available on https://github.com/subugoe/oauni.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abediyarandi, N., & Mayr, P. (2019). The state of open access in Germany: An analysis of the publication output of German universities. In *Proceedings of the 17th international conference on scientometrics & informetrics (ISSI 2019)* (pp. 2768–2769). https://arxiv.org/abs/1905.00011.

AG Ressortforschungseinrichtungen. (2013). *Forschen—Prüfen—Beraten. Ressortforschungseinrichtungen als Dienstleister für Politik und Gesellschaft. Positionspapier der Arbeitsgemeinschaft der Ressortforschungseinrichtungen.* https://www.ressortforschung.de/de/res_medien/fpb_positionspapier.pdf.

Akbaritabar, A., & Stahlschmidt, S. (2019). Applying crossref and unpaywall information to identify gold, hidden gold, hybrid and delayed Open Access publications in the KB publication corpus. *SocArXiv.* https://doi.org/10.31235/osf.io/sdzft.

Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). Proportion of open access papers published in peer-reviewed journals at the European and World levels—1996–2013. *Copyright, Fair Use, Scholarly Communication, Etc.* https://digitalcommons.unl.edu/scholcom/8.

Barlösius, E. (2010). Ressortforschung. In D. Simon, A. Knie, & S. Hornbostel (Eds.), *Handbuch Wissenschaftspolitik* (pp. 377–389). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91993-5_26.

Bosman, J., & Kramer, B. (2018). Open access levels: A quantitative exploration using Web of Science and oaDOI data. *PeerJ Preprints* 6:e3520v1. https://doi.org/10.7287/peerj.preprints.3520v1.

Bruch, C., Deinzer, G., Geschuhn, K., Hätscher, P., Hillenkoetter, K., Kress, U., Pampel, H., Schäffler, H., Stanek, U., Timm, A., Wagner, A., Gebert, A., Hanig, K., Herbstritt, M., Mruck, K., Scheiner, A., Scholze, F., Schulze, M., Siegert, O., & Vierkant, P. (2015). *Positions on creating an Open Access publication market which is scholarly adequate: Positions of the Ad Hoc Working Group Open Access Gold in the priority initiative 'Digital Information' of the Alliance of Science Organisations in Germany. Ad-hoc-Arbeitsgruppe Open-Access-Gold der Schwerpunktinitiative 'Digitale Information' der Allianz der deutschen Wissenschaftsorganisationen.* https://doi.org/10.2312/allianzoa.009.

Bruns, A., Lenke, C., Schmidt, C., & Taubert, N. (2019). ISSN-matching of gold OA Journals (ISSN-GOLD-OA) 3.0. Bielefeld University. https://doi.org/10.4119/unibi/2934907.

Bundesministerium für Bildung und Forschung (BMBF). (2016). Organisationen und Einrichtungen in Forschung und Wissenschaft. *Bundesberichte Forschung und Innovation. Ergänzungsband II*.

Chamberlain, S., Zhu, H., Jahn, N., Boettiger, C., & Ram, K. (2020). rcrossref: Client for various 'CrossRef' 'APIs' *(Version R package version 1.0.0)* [Computer software]. https://CRAN.R-project.org/package=rcrossref.

Donner, P., Rimmert, C., & van Eck, N. J. (2020). Comparing institutional-level bibliometric research performance indicator values based on different affiliation disambiguation systems. *Quantitative Science Studies, 1*(1), 150–170. https://doi.org/10.1162/qss_a_00013

Dusdal, J., Powell, J. J. W., Baker, D. P., Fu, Y. C., Shamekhi, Y., & Stock, M. (2020). University vs. research institute? The dual pillars of German science production, 1950–2010. *Minerva*. https://doi.org/10.1007/s11024-019-09393-2.

Else, H. (2018). Dutch publishing giant cuts off researchers in Germany and Sweden. *Nature, 559*(7715), 454–455. https://doi.org/10.1038/d41586-018-05754-1

Eppelin, A., Pampel, H., Bandilla, W., & Kaczmirek, L. (2012). Umgang mit Open-Access-Publikationsgebühren – die Situation in Deutschland in 2010. *GMS Medizin - Bibliothek - Information; 12(1–2):Doc04.* https://doi.org/10.3205/MBI000240.

Fournier, J. (2007). Open Access in der Deutschen Forschungsgemeinschaft. Positionen, Projekte, Perspektiven. *Zeitschrift für Bibliothekswesen und Bibliographie*, *54*(4/5), 224–229. https://doi.org/10.3196/18642950085445130.

Fournier, J., & Weihberg, R. (2013). Das Förderprogramm »Open Access Publizieren« der Deutschen Forschungsgemeinschaft. Zum Aufbau von Publikationsfonds an wissenschaftlichen Hochschulen in Deutschland. *Zeitschrift für Bibliothekswesen und Bibliographie*, *60*(5), 236–243. https://doi.org/10.3196/186429501360528.

Goebelbecker, J. (2005). The role of publications in the new programme oriented funding of the Hermann von Helmholtz Association of National Research Centres (HGF). *Scientometrics, 62*(1), 173–181. https://doi.org/10.1007/s11192-005-0012-x

Huang, C.-K., Neylon, C., Hosking, R., Montgomery, L., Wilson, K. S., Ozaygen, A., & Brookes-Kenworthy, C. (2020). Evaluating the impact of open access policies on research institutions. *eLife, 9*, e57067. https://doi.org/10.7554/eLife.57067

Jahn, N., & Tullney, M. (2016). A study of institutional spending on open access publication fees in Germany. *PeerJ, 4*, e2323. https://doi.org/10.7717/peerj.2323

Kohls, A., & Mele, S. (2018). Converting the literature of a scientific field to open access through global collaboration: The experience of SCOAP3 in Particle Physics. *Publications, 6*(2), 15. https://doi.org/10.3390/publications6020015

Laakso, M., & Björk, B.-C. (2012). Anatomy of open access publishing: A study of longitudinal development and internal structure. *BMC Medicine, 10*(1), 124. https://doi.org/10.1186/1741-7015-10-124

Lagoze, C., & Van de Sompel, H. (2003). The making of the open archives initiative protocol for metadata harvesting. *Library Hi Tech, 21*(2), 118–128. https://doi.org/10.1108/07378830310479776

Larivière, V., & Sugimoto, C. R. (2018). Do authors comply when funders enforce open access to research? *Nature, 562*(7728), 483–486. https://doi.org/10.1038/d41586-018-07101-w

Lex, A., & Gehlenborg, N. (2014). Sets and intersections. *Nature Methods, 11*(8), 779–779. https://doi.org/10.1038/nmeth.3033

Lossau, N., & Peters, D. (2008). DRIVER: Building a sustainable infrastructure of European scientific repositories. *LIBER Quarterly, 18*(3–4), 437. https://doi.org/10.18352/lq.7942

Martín-Martín, A., Costas, R., van Leeuwen, T., & Delgado López-Cózar, E. (2018). Evidence of open access of scientific publications in Google Scholar: A large-scale analysis. *Journal of Informetrics, 12*(3), 819–841. https://doi.org/10.1016/j.joi.2018.06.012

Mitchell, A. D. (1998). *The Fraunhofer Society: A unique German contract research organization comes to America.* The Office.

Mueller-Langer, F., & Watt, R. (2018). How many more cites is a $3,000 open access fee buying you? Empirical evidence from a natural experiment. *Economic Inquiry, 56*(2), 931–954. https://doi.org/10.1111/ecin.12545

Müller, U., & Schirmbacher, P. (2007). Der 'grüne Weg zu Open Access' in Deutschland. *Zeitschrift für Bibliothekswesen und Bibliographie, 54*(4/5), 183–193. https://doi.org/10.3196/1864295008544570

National Science Board, National Science Foundation. (2019). Publication output: U.S. trends and international comparisons. *NSB-2020-6; Science and Engineering Indicators 2020*. https://ncses.nsf.gov/pubs/nsb20206/.

Olsson, L., Francke, H., Lindelöw, C. H., & Willén, N. (2020). The first Swedish Read & Publish agreement: An evaluation of the Springer Compact pilot. *LIBER Quarterly, 30*(1), 1. https://doi.org/10.18352/lq.10309

Peacock, V. (2016). Academic precarity as hierarchical dependence in the Max Planck Society. *HAU: Journal of Ethnographic Theory, 6*(1), 95–119. https://doi.org/10.4318/hau6.1.006

Pieper, D., & Summann, F. (2006). Bielefeld Academic Search Engine (BASE): An end-user oriented institutional repository search service. *Library Hi Tech, 24*(4), 614–619. https://doi.org/10.1108/07378830610715473

Pinfield, S. (2015). Making open access work: The "state-of-the-art" in providing open access to scholarly literature. *Online Information Review, 39*(5), 604–636. https://doi.org/10.1108/OIR-05-2015-0167

Pinfield, S., Salter, J., & Bath, P. A. (2016). The "total cost of publication" in a hybrid open-access environment: Institutional approaches to funding journal article-processing charges in combination with subscriptions. *Journal of the Association for Information Science and Technology, 67*(7), 1751–1766. https://doi.org/10.1002/asi.23446

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ, 6*, e4375. https://doi.org/10.7717/peerj.4375

Piwowar, H., Priem, J., & Orr, R. (2019). The future of OA: A large-scale analysis projecting open access publication and readership [Preprint]. bioRxiv. https://doi.org/10.1101/795310.

Pölönen, J., Laakso, M., Guns, R., Kulczycki, E., & Sivertsen, G. (2020). Open access at the national level: A comprehensive analysis of publications by Finnish researchers. *Quantitative Science Studies*. https://doi.org/10.1162/qss_a_00084

Powell, J. J. W., & Dusdal, J. (2017). Science production in Germany, France, Belgium, and Luxembourg: Comparing the contributions of research universities and institutes to Science, Technology, Engineering, Mathematics, and Health. *Minerva, 55*(4), 413–434. https://doi.org/10.1007/s11024-017-9327-z

Rimmert, C., Schwechheimer, H., & Winterhager, M. (2017). *Disambiguation of author addresses in bibliometric databases—Technical report.* [Report]. https://pub.uni-bielefeld.de/record/2914944.

Robinson-Garcia, N., Costas, R., & van Leeuwen, T. N. (2020). Open Access uptake by universities worldwide. *PeerJ, 8*, e9410. https://doi.org/10.7717/peerj.9410

Rovira, A., Urbano, C., & Abadal, E. (2019). Open access availability of Catalonia research output: Case analysis of the CERCA institution, 2011–2015. *PLoS ONE, 14*(5), e0216597. https://doi.org/10.1371/journal.pone.0216597

Schimank, U. (2005). 'New Public Management' and the academic profession: Reflections on the German situation. *Minerva, 43*(4), 361–376. https://doi.org/10.1007/s11024-005-2472-9

Schimmer, R., Geschuhn, K., & Palzenberger, M. (2013). Open Access in Zahlen: Der Umbruch in der Wissenschaftskommunikation als Herausforderung für Bibliotheken. *Zeitschrift für Bibliothekswesen und Bibliographie, 60*(5), 244–250. https://doi.org/10.3196/186429501360532

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.

Schirrwagen, J., Manghi, P., Manola, N., Bolikowski, L., Rettberg, N., & Schmidt, B. (2013). Data curation in the OpenAIRE Scholarly communication infrastructure. *Information Standards Quarterly, 25*(3), 13. https://doi.org/10.3789/isqv25no3.2013.03

Schmidt, B., & Ilg-Hartbecke, K. (2009). Open Access am Standort D – erweiterte Perspektiven für die Wissenschaft. *GMS Medizin - Bibliothek - Information*, *9*(1), Doc05. https://doi.org/10.3205/mbi000133.

Severin, A., Egger, M., Eve, M. P., & Hürlimann, D. (2020). Discipline-specific open access publishing practices and barriers to change: An evidence-based review. *F1000Research, 7*, 1925. https://doi.org/10.2688/f1000research.17328.2

Sikora, A., & Geschuhn, K. (2015). Management of article processing charges—Challenges for libraries. *Insights the UKSG Journal, 28*(2), 87–92. https://doi.org/10.1629/uksg.229

Spezi, V., Wakeling, S., Pinfield, S., Creaser, C., Fry, J., & Willett, P. (2017). Open-access mega-journals: The future of scholarly communication or academic dumping ground? A review. *Journal of Documentation, 73*(2), 263–283. https://doi.org/10.1108/JD-06-2016-0082

Stahlschmidt, S., Stephen, D., & Hinze, S. (2019). *Performance and structures of the German science system* (No. 5–2019; Studien Zum Deutschen Innovationssystem). http://hdl.handle.net/10419/194275.

Statistisches Bundesamt. (2019). Bildung und Kultur. Personal an Hochschulen. *Fachserie 11,* Reihe 4.4 (pp. 1–367).

Statistisches Bundesamt. (2020a). Finanzen und Steuern. Ausgaben, Einnahmen und Personal der öffentlichen und öffentlich geförderten Einrichtungen für Wissenschaft. *Fachserie 14*, Reihe 3.6; p. 93. Statistisches Bundesamt. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Forschung-Entwicklung/Publikationen/Downloads-Forschung-Entwicklung/ausgaben-einnahmen-personal-2140360187004.pdf.

Statistisches Bundesamt. (2020b). Bildung und Kultur. Finanzen an Hochschulen. *Fachserie 11, Reihe 4.5,* pp. 1–212.

Suber, P. (2012). Open Access. MIT Press. https://dash.harvard.edu/handle/1/10752204.

Taubert, N. C. (2019). *Fremde Galaxien und abstrakte Welten—Open Access in Astronomie und Mathematik: Eine soziologische Analyse*. Bielefeld: Transcript.

Van der Graaf, M. (2017). Paying for open access: The author's perspective. Zenodo. https://doi.org/10.5281/ZENODO.438037

Vogel, G. (2019). More than 700 German research institutions strike open-access deal with Springer Nature. *Science*. https://doi.org/10.1126/science.aaz2308

Voigt, M., Winterhalter, C., Riesenweber, C., & Hübner, A. (2018). *Open-Access-Anteil bei Zeitschriftenartikeln von Wissenschaftlerinnen und Wissenschaftlern an Einrichtungen des Landes Berlin: Datenauswertung für das Jahr 2016* [Report]. https://depositonce.tu-berlin.de/handle/11303/7682

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wissenschaftsrat. (2013). *Perspektiven des deutschen Wissenschaftssystems*. https://www.wissenschaftsrat.de/download/archiv/3228-13.pdf.

Wohlgemuth, M., Rimmert, C., & Taubert, N. C. (2017). *Publikationen in Gold-Open-Access-Journalen auf globaler und europäischer Ebene sowie in Forschungsorganisationen*. Bielefeld University. https://pub.uni-bielefeld.de/record/2912807.