Initiating discipline-specific Open Science Monitoring with the Open Science Dashboard for Earth Sciences

Maaike Duine[*], Anastasiia Iarkaeva[**] and Andreas Hübner[***]

[*] *maaike.duine@open-access-berlin.de*
https://orcid.org/0000-0003-3412-7192
Open-Access-Büro Berlin, Universitätsbibliothek der Freien Universität Berlin, Germany

[**] *anastasiia.iarkaeva@charite.de*
https://orcid.org/0000-0002-7370-1663
Berlin Institute of Health at Charité (BIH), QUEST Center for Responsible Research, Germany

*** *andreas.huebner@fu-berlin.de*
https://orcid.org/0000-0001-7342-9789
Forschungs- und Publikationsservices, Universitätsbibliothek der Freien Universität Berlin, Germany

Abstract
With the increased importance attributed to Open Science practices, Open Science Monitoring becomes more relevant as well. As Open Science practices differ across scientific disciplines, monitoring models should be flexible and context should always be provided. In this paper, we describe how we initiated discipline-specific monitoring with the development of an Open Science Dashboard for the Department of Earth Sciences at Freie Universität Berlin.

Keywords: Open Science Monitoring; Open Science Indicators; Open Access; Persistent Identifiers

1. Introduction

1.1. Diversity in Open Science practices across disciplines and communities

Open Science (OS) or Open Research is increasingly relevant in scientific life, aiming to make the entire research process, its sources and results openly accessible and reusable in the long term. OS embraces many principles and practices such as Open Access (OA) publishing, Open Data, Open Peer Review, Open Educational Resources and Citizen Science (*UNESCO Recommendation on Open Science, 2021*). According to the UNESCO definition, one of the guiding principles for implementing OS is 'flexibility': to acknowledge that there is no one-size-fits-all way to practise OS. The difference in implementation and uptake of OS practices across various scientific disciplines is confirmed by multiple research studies: e.g. for OA article publishing (Severin et al., 2020), Citizen Science practices (Pettibone et al., 2017), preprint publishing (Chiarelli et al., 2019; Fry et al., 2016), and data sharing practices (Tedersoo et al., 2021; Zuiderwijk & Spiers, 2019; Khan et al., 2023).

As OS practices gain importance, OS monitoring is becoming more relevant as well. Monitoring assesses developments in OS, enables broader insight into OS trends over time, and potentially exposes training and infrastructural needs. However, OS monitoring mainly focuses on OA journal articles (Himanen & Nykyri, 2024). In order to recognise, evaluate and credit much more diverse open research outputs, there is a need for contextualised and discipline-specific monitoring. Global initiatives to reform research assessment, like the San

Francisco Declaration on Research Assessment[1] and the Coalition for Advancing Research Assessment Agreement[2], also call for rewarding more diverse research outputs.

1.2. BUA Open Science Dashboards and Open Science Magnifiers Projects

The Berlin University Alliance (BUA) is composed of four institutions: Freie Universität Berlin, Humboldt-Universität zu Berlin, Technische Universität Berlin, Charité – Universitätsmedizin Berlin. The BUA Open Science Dashboards project was initiated in 2021 (followed by the implementation project BUA OS Magnifiers in 2024) (Bobrov et al., 2024), by project partners QUEST Center for Responsible Research and Open-Access-Büro Berlin. One of the project's aims is to collaborate with scientific communities to develop discipline-specific OS indicators and metrics, and visualise the collected metrics in dashboards to support OS monitoring.

An OS dashboard gives a visual overview of OS metrics, offering the advantage of monitoring trends over time and gaining insights through interactive graphs and filters. Within the biomedical community, a dashboard had already been established by QUEST: the Charité Dashboard on Responsible Research[3]. This dashboard encompasses several OS metrics, such as OA publishing, Open Data, Open Code, Preprints, or more discipline-specific - Clinical Trials metrics.

1.3. Initiation of OS monitoring in the Earth Sciences

OS is generally highly valued by researchers at Berlin institutions (Lüdtke & Ambrasat, 2022). Freie Universität Berlin (FU Berlin) supports OS through the BUA Mission Statement for OS[4], and the university is subject to the OS provisions of the Berlin Higher Education Act (*Gesetz über die Hochschulen im Land Berlin*, 2021). The university does not monitor any OS metrics individually, but has several OS-related policies (*OA Policy of FU Berlin*, 2021; *Research Data Policy of FU Berlin*, 2021) in place. The university library offers OS-related workshops and consultations with researchers.

The Department of Earth Sciences at FU Berlin collaborates with the University Library in several OS-related activities, including joint data management projects in the NFDI4Earth consortium to increase visibility and quality of the department's research data publications. Additionally, library staff promotes good scientific practice and openness in student seminars.

Initial discussions with the Institute of Geographical Sciences of the Department of Earth Sciences at FU Berlin revealed that OS practices are adopted but also vary *within* the Earth Sciences. The institute encompasses various fields of study, (e.g., human and (applied) physical geography) each with unique publishing practices and distinct research methods, such as data collection. This diversity results in various types of research outputs and, consequently, presents different challenges for openness and preservation.

[1] https://sfdora.org/

[2] https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

[3] https://quest-dashboard.charite.de/#tabStart

[4] https://www.berlin-university-alliance.de/en/commitments/research-quality/open-science/Leitbild-fuer-OS/index.html

Earth Sciences is one of the research fields in which OA journal article publishing is comparatively widespread (Severin et al., 2020). However, more inclusive metrics, beyond journal article publications, are needed. Pourret et al. (2022) claim that monitoring and recognising open publishing of preprints, datasets, engagement and communication with the public, are important, especially in research intersecting with the environment, climate change, and biodiversity. As there was, to our knowledge, no existing dashboard for OS monitoring purposes in the Earth Sciences we investigated the following question: *Which OS metrics can be collected and visualised in a dashboard for the Earth Sciences, taking into account the specific publishing culture in this discipline?*

2. Methods

2.1. Data collection

We started the collection of OS metrics with OA publications, analysing the Department of Earth Sciences' list of publications (2016–2022) from the FU Berlin university bibliography (UB)[5]. The Department of Earth Sciences comprises three institutes: Institute of Geological Sciences, Institute of Geographical Sciences and Institute of Meteorology. The bibliography is considered the most comprehensive source of "classic" research outputs because of its primary function of supporting the performance-based funding process at FU Berlin.

2.2 Data enrichment

The UB publication list underwent deduplication and data enrichment processes, and was then sorted according to the three institutes. Publications were categorised into eligible types, including "journal article", "book", "book chapter", "conference paper", "conference abstract", and "other research outputs". The latter category includes, e.g., book reviews, project reports, book chapters in school books, or electronic supplementary material. The classification of these output types as "other" was based on their relative rarity. All publications were manually recategorised due to variations in categorisation within the FU UB; e.g., the category 'Contribution to a book' could be either abstract or book chapter.

Journal articles are the most common publication type (62.0%) within the Department of Earth Sciences. Consequently, they were analysed separately. The remaining publication types are referred to as "non-journal-article output".

We identified and assigned Persistent Identifiers (PIDs), including DOIs, where possible, and checked the OA and copyright/licence status. Entries without a PID underwent individual research, with an average of 5 minutes spent per output. If no PID was found, the URL was included.

For entries with DOIs provided by Crossref, their OA statuses and categories were cross-referenced using the Unpaywall REST API[6]. Entries lacking Crossref DOIs could not be assigned an OA status by Unpaywall. Furthermore, for validated DOIs, many assigned OA statuses and categories through Unpaywall were inaccurate. This necessitated a manual verification and reassignment for all entries, particularly for non-journal-article output. Published research datasets in the UB list were excluded from this analysis.

[5] https://frub-berlin.primo.exlibrisgroup.com/

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

The OA status and category determination rules are documented in Duine et al. (2024). Categories: “gold”, “green”, and “hybrid” categories are OA; “bronze” and “closed” are closed. The category “bronze” is treated as closed access, as these articles lack licence information and thus cannot be openly reused (Severin et al., 2020). Additionally, types of open licences and publisher names were collected. Table 1 summarises the corrected and enriched publication types during manual validation.

Table 1. Data enrichment of the UB data through PID and OA status verification.

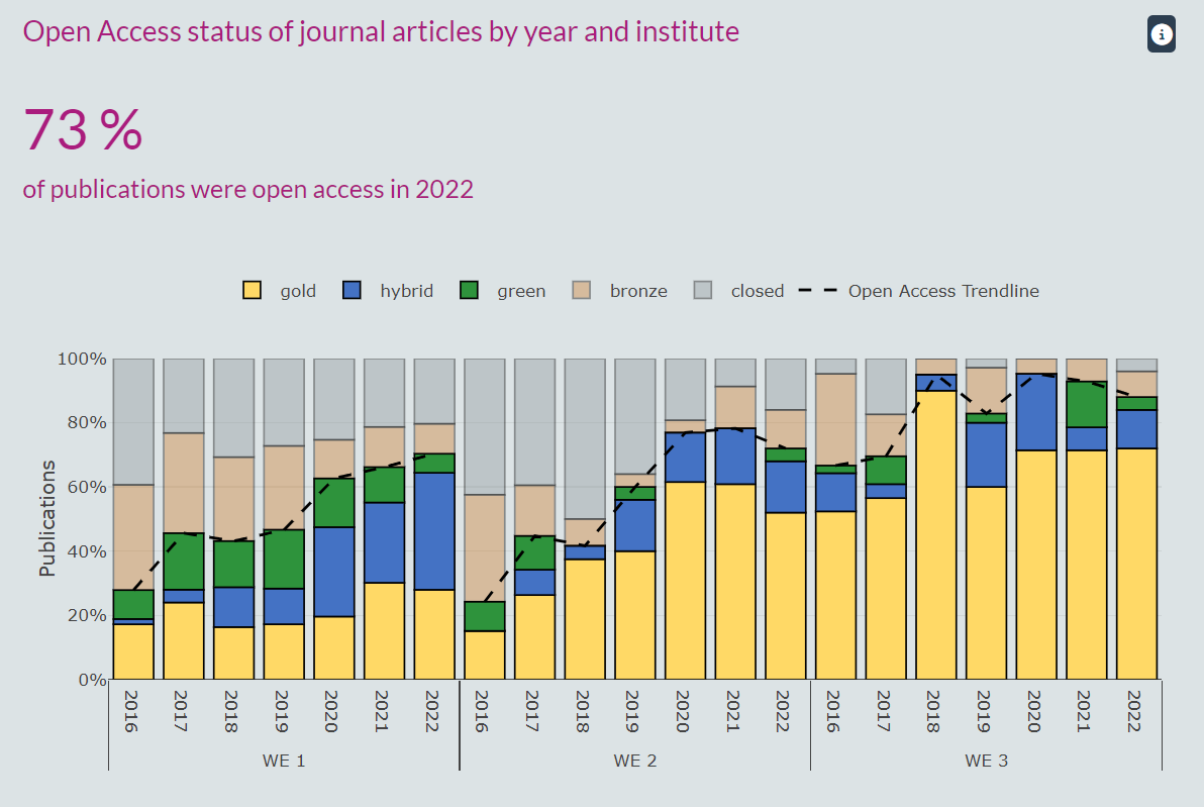
Publication type	**Total**	**Initially with PID (% from total in category)**	**With PID after data enrichment (% from total in category)**	**Publications with OA status information that were initially correct (% from total in category)**	**Publications with correct OA status information after data enrichment (% from total in category)**	**Publications with OA status information and PID information that were initially correct (% from total in category)**
Journal article	1366	1169 (85.6%)	1366 (100%)	616 (45.1%)	1366 (100%)	536 (39.2%)
Book	26	8 (30.1%)	17 (65.4%)	1 (3.8%)	26 (100%)	1 (3.8%)
Book chapter	113	36 (31.9%)	69 (61.1%)	0 (0%)	113 (100%)	0 (0%)
Conference paper	172	17 (9.9%)	21 (12.2%)	1 (0.6%)	172 (100%)	0 (0%)
Conference abstract	437	123 (28.1%)	160 (36.6%)	1 (0.2%)	437 (100%)	0 (0%)
Other	88	13 (14.8%)	21 (23.9%)	4 (4.5%)	88 (100%)	3 (3.4%)
Total	2202	1366 (62.0%)	1654 (75.1%)	623 (28.3%)	2202 (100%)	539 (24.6%)

2.3 Graphical representation in the dashboard

The dashboard[7] currently displays the OA categories for all publication types, along with Creative Commons (CC) licences for journal articles. For non-journal-article output, it also displays PID availability to highlight the importance of PIDs for linking research content as well as OA categories.

The dashboard was developed using Shiny, an R package for interactive web applications[8]. It features bar plots and interactive components, such as sliders and checkboxes, allowing users to explore the data dynamically. User inputs through filters generate reactive outputs/plots that enable users to examine the results of individual institutes in both relative and absolute numbers, illustrating developments over time.

Figure 1: OA monitoring of journal articles through the interactive dashboard (exemplary graph); WE 1 = Institute of Geological Sciences, WE 2 = Institute of Geographical Sciences, WE 3 = Institute of Meteorology.



An overview of data collection and visualisation methods is provided with the dashboard. The underlying data and code is openly shared. Ongoing maintenance and updates are conducted to ensure the dashboard's functionality and relevance. Data preservation on Zenodo enables reproducibility. Code preservation on GitHub enables version control and reproducibility.

[7] https://quest-open-earthsciences.charite.de/

[8] https://www.rstudio.com/products/shiny/

3. Results

3.1 OA status and category of journal articles

A total of 1366 journal articles from 2016-2022 were analysed (see Table 2). On average, the share of OA publications in 2016-2022 across all institutes was 56.1%. Figure 1 shows the OA category distribution over 2016-2022, highlighting data from 2022. Notably, the Institute of Meteorology showed a substantially higher OA share 81.7%) than other institutes.

Table 2. OA distribution in journal articles.

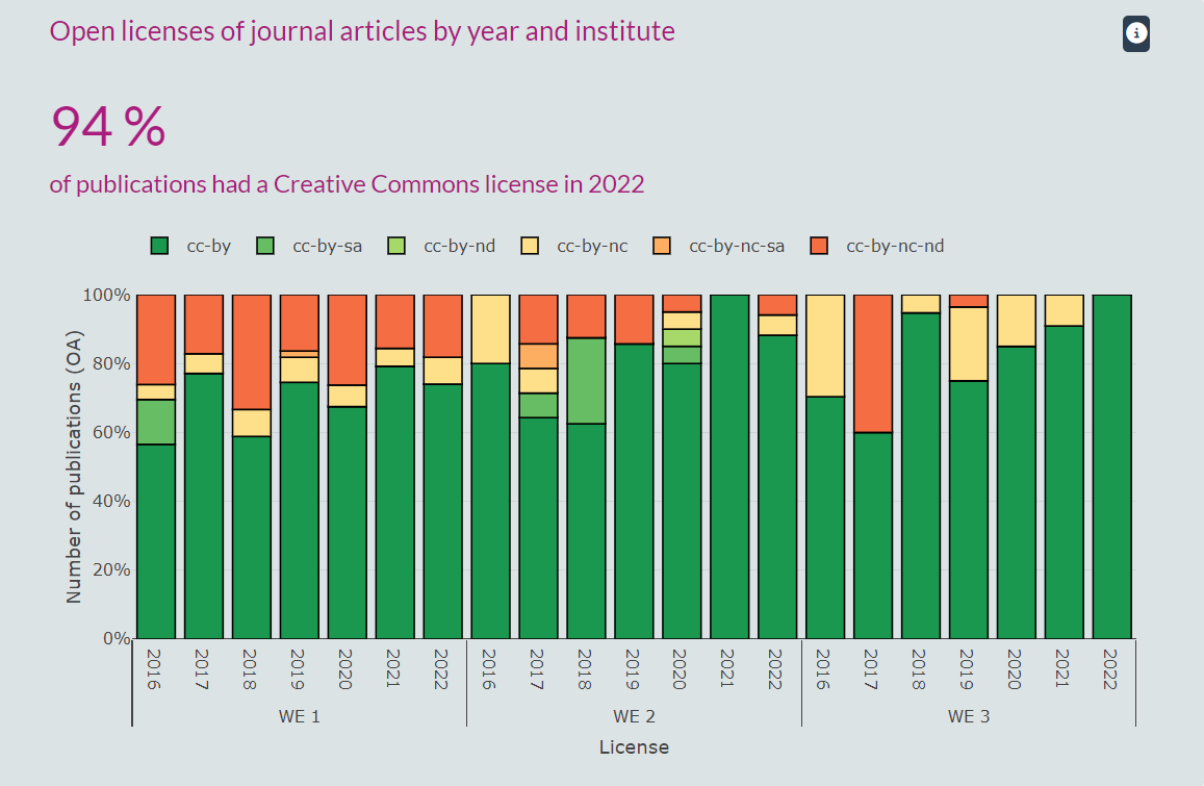
Institute name	**Total**	**Bronze**	**Closed**	**Gold**	**Green**	**Hybrid**
Institute of Geological Sciences (WE 1)	992	213	266	212	134	167
Institute of Geographical Sciences (WE 2)	194	27	61	77	9	20
Institute of Meteorology (WE 3)	180	25	8	117	7	23
All institutes	1366	265	335	406	150	210

Manual verification of OA statuses revealed that 54.9% of the original OA statuses in the UB did not match the manually validated statuses (based on binary OA and non-OA status). Moreover, 42 journal articles could not be automatically validated via Unpaywall. Notably, we compared manually validated OA statuses and categories with results from Unpaywall, both coded in binary format, and found that 46.2% did not match.

3.2. Open licences of journal articles

82.9% of all OA journal articles in 2016-2022 are published with CC licences, with the CC BY licence being most prevalent (62.2%). In 2022, 94% of OA articles were published under CC licences, maintaining a high level similar to the overall 84.9% between 2016-2022.

Figure 2: Creative Commons licences of journal articles in the OS Dashboard; WE 1 = Institute of Geological Sciences, WE 2 = Institute of Geographical Sciences, WE 3 = Institute of Meteorology.



3.3. Non-journal-article outputs

A total of 836 non-journal-article outputs were analysed (Institute of Geological Sciences: 608, Institute of Geographical Sciences: 107, Institute of Meteorology: 121). 35.8% of these are openly accessible, and 34.4% have DOIs or Handle identifiers. The large number of 836 non-journal-article outputs in our sample represent a significant scientific communication venue. However, a significant portion of research outputs categorised as "book chapter", "conference paper", and "conference abstract" (> 85% of all non-journal-article outputs) lack PIDs.

4. Discussion

4.1. Data source / data enrichment

Although the FU UB is considered the most comprehensive source of publication information, extensive manual research was required to improve data quality. While PIDs for journal articles were mostly covered, the quality of OA statuses and categories was poor, and licence information was lacking. We queried Web of Science (WoS) and OpenAlex to explore automatic data information enrichment.

A discipline-specific search using WoS yielded 623 journal publications (WoS subject category=Geosciences, Multidisciplinary; Publication year=2016-2022; at least one author from FU). 65% (408) of articles are classified as OA. While OA category information, including distinctions such as "gold", "green" etc., is available, licence information is missing, and the database does not support analysis of individual university departments.

A search in the OpenAlex database (Institution=FU Berlin; Field=Earth and Planetary Sciences; Year=2016-2022; date of search: 26 March 2024) yielded 1320 results, with 1250 classified as “articles”. However, only 464 results (457 of which are journal articles) matched DOIs in the FU UB.

This discrepancy primarily arises because OpenAlex allows filtering by “Earth and Planetary Sciences", but not specifically by “Department of Earth Sciences of FU Berlin”. Consequently, many publications in the “Earth and Planetary Sciences” field in OpenAlex originate from co-authors affiliated with other departments at FU Berlin (e.g., DOI[9], while some publications from authors of the Department of Earth Sciences at FU Berlin are not categorised in this OpenAlex field (e.g., DOI[10]).

This comparison with other databases shows that merging UB data with other sources presents significant challenges. Automated mapping of OA status and category requires structured and comparable data. Challenges encountered in mapping publication data with Unpaywall have been described by Stricker (2023). Additionally, databases like WoS or OpenAlex lack the granularity for departmental or institutional-level analysis.

We conclude that the UB data quality is currently insufficient for our OS monitoring approach. While the manual data enrichment efforts produced the highest quality results possible at this point, they were resource-intensive and not scalable. Therefore, using FU UB data as a source for monitoring OA uptake across more departments or the entire university would require substantial resources for both dataset enrichment and refinement. Using the UB data (or databases like WoS) “as is” may offer scalability for larger publication corpora, but data quality and coverage of publication types are severely compromised.

4.2. OS metrics in the current OS Dashboard Earth Sciences

In 2023 and 2024, the dashboard was introduced at the three institutes of the Department of Earth Sciences. It was positively acknowledged by the researchers that OA monitoring enables department or research group leadership to identify gaps in OA publishing activities. Additionally, the dashboard encourages more general discussions on OS or potentially identifies training needs. PID availability for non-journal-article outputs was discussed controversially, as lack of PIDs can lead to incorrect citations, limited reach to relevant target communities, and exclusion from monitoring. Moreover, PIDs contribute to making research entities more FAIR (Findable, Accessible, Interoperable, and Reusable), facilitating integration into broader scholarly networks (Cousijn et.al., 2021).

5. Conclusions and next steps

Relying solely on available, non-enriched data for OS monitoring excludes relevant research outputs like conference abstracts, and overlooks OS practices like applying CC licences. We acknowledge that our monitoring approach is one of many, and that manual data enrichment is impractical for large datasets. However, data sources like OpenAlex are not as comprehensive - yet - for our monitoring approach. Salamoura and Tsakonas (2024) already pointed out that

[9] 10.5194/isprs-archives-XLVI-M-1-2021-531-2021
[10] 10.1073/pnas.1714341115

monitoring models vary significantly in terms of data sources, information representation, and granularity, and that one of the key improvements needed is the establishment of clear and concise terminology. We recommend that all relevant stakeholders, such as disciplinary communities, infrastructure providers, publishers, conference organisers, and repository managers, continue to collaborate on implementing PIDs and reliable, structured metadata to enable scalable, high-quality screening, including more automated processes.

We are currently working on the inclusion of Open Data and Open Code sharing as a further OS metric in the dashboard. We also aim to explore whether additional OS metrics like ORCID iDs, preprint publishing and open dissertations could be included. As digital curation of physical samples is increasingly important in the Earth Sciences, inclusion of the International Generic Sample Number (IGSN) will be tested. We will continue to explore possible expansions of the dashboard, also taking into account recent calls for a more systemic monitoring framework that not only focuses on outputs but also on processes, outcomes and impacts (Rafols et.al., 2024).

Open science practices
The FU bibliography has been retrieved from https://frub-berlin.primo.exlibrisgroup.com/ and is openly available. Data curated during this study is openly available on Zenodo (https://doi.org/10.5281/zenodo.13969494). We have made the source code for our project available on GitHub at https://github.com/quest-bih/open-earthsciences.

Acknowledgments
The authors would like to thank Tobias Tilger and Kai Nolte for extensive data enrichment, Dr. Evgeny Bobrov and Dr. Maxi Kindling for their extensive comments to earlier versions of this paper, and two anonymous reviewers for their valuable feedback.

Author contributions
Maaike Duine: Conceptualization, Methodology, Writing – original draft, Writing – review & editing; Anastasiia Iarkaeva: Conceptualization, Methodology, Software, Formal Analysis, Visualization, Data curation, Writing – original draft, Writing – review & editing; Andreas Hübner: Methodology, Writing – original draft, Writing – review & editing

Competing interests
The authors have no competing interests to declare.

Funding information
The BUA OS Dashboards project and the BUA OS Magnifiers project are funded by the Federal Ministry of Education and Research (BMBF) and the state of Berlin under the Excellence Strategy of the Federal Government and the Länder through the Berlin University Alliance.

References

Bobrov, E., Duine, M., Kindling, M. & Iarkaeva, A. (2024). Das Projekt BUA Open Science Magnifiers: Weiterentwicklung des Open-Science-Monitorings in verschiedenen Disziplinen und Forschungsbereichen. *Open Access Blog Berlin*. https://doi.org/10.59350/yp5f9-wjn07

Chiarelli, A., Johnson, R., Pinfield, S. & Richens, E. (2019). Preprints and Scholarly Communication: An Exploratory Qualitative Study of Adoption, Practices, Drivers and Barriers [version 2; peer review: 3 approved, 1 approved with reservations]. *F1000Research 2019*, 8:971. https://doi.org/10.12688/f1000research.19619.2

Cousijn, H., Braukmann, R., Fenner, M., Ferguson, C., Van Horik, R., Lammey, R., Meadows, A. & Lambert, S. (2021). Connected Research: The potential of the PID Graph. *Patterns*, 2(1), 100180. https://doi.org/10.1016/j.patter.2020.100180

Duine, M., Hübner, A. & Iarkaeva, A. (2024). Enrichment of university bibliography data for open science monitoring. Zenodo. https://doi.org/10.5281/zenodo.10998219

Fry, J., Spezi, V., Probets, S., & Creaser, C. (2016). Towards an understanding of the relationship between disciplinary research cultures and open access repository behaviors. *Journal of the Association for Information Science and Technology*, 67(11), 2710–2724. https://doi.org/10.1002/asi.23621

Gesetz über die Hochschulen im Land Berlin (Berliner Hochschulgesetz - BerlHG), § 41 - Forschungsberichte. (2021). https://gesetze.berlin.de/bsbe/document/jlr-HSchulGBE2011V27P41

Himanen, L. & Nykyri, S. (2024). Towards a sustainable and responsible model for monitoring open science and research—analysis of the Finnish model for monitoring open science and research. *Research Evaluation*, rvae008. https://doi.org/10.1093/reseval/rvae008

Iarkaeva, A., Duine, M., & Hübner, A. (2024). Dataset for Earth Sciences at Freie Universität Berlin: Open Access, Licenses and Persistent Identifiers Monitoring (Version v2) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13969494

Khan, N., Thelwall, M. & Kousha, K. (2023). Data sharing and reuse practices: Disciplinary differences and improvements needed. *Online Information Review*, 47(6), 1036–1064. https://doi.org/10.1108/OIR-08-2021-0423

Lüdtke, D. & Ambrasat, J. (2022). Basic Report Berlin Science Survey. *edoc-Server*. https://doi.org/10.18452/26223

Open Access Policy of Freie Universität Berlin. (2021). http://dx.doi.org/10.17169/refubium-31442

Pettibone, L., Vohland, K. & Ziegler, D. (2017). Understanding the (inter)disciplinary and institutional diversity of citizen science: A survey of current practice in Germany and Austria. *PLOS ONE*, 12(6), e0178778. https://doi.org/10.1371/journal.pone.0178778

Pourret, O., Irawan D.E., Shaghaei N., van Rijsingen E.M. & Besançon L. (2022). Toward more inclusive metrics and open science to measure research assessment in earth and natural sciences. *Frontiers in Research Metrics and Analytics*, 7. https://doi.org/10.3389/frma.2022.850333

Rafols, I., Meijer I., and Molas-Gallart J. (2024). Monitoring Open Science as transformative change: Towards a systemic framework [version 1; peer review: 2 approved]. *F1000Research*, 13:320. https://doi.org/10.12688/f1000research.148290.1

Research Data Policy of Freie Universität Berlin. (2021). http://dx.doi.org/10.17169/refubium-32141

Salamoura, A. & Tsakonas, G. (2024). On the challenges of open access monitoring. *Insights*, 37(1). https://doi.org/10.1629/uksg.641

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/10.12688/f1000research.17328.2

Stricker, M. (2023). Open Access Monitoring: Verzerrende Datenquellen und unbeabsichtigte Leerstellen – eine explorative Studie. *Bibliothek Forschung und Praxis,* 47(2), 382-392. https://doi.org/10.1515/bfp-2023-0015

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K. & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1). https://doi.org/10.1038/s41597-021-00981-0

UNESCO Recommendation on Open Science. (2021). https://doi.org/10.54677/mnmh854

Zuiderwijk, A. & Spiers, H. (2019). Sharing and re-using open data: A case study of motivations in astrophysics. *International Journal of Information Management*, 49, 228–241. https://doi.org/10.1016/j.ijinfomgt.2019.05.024