Learned Publishing

LEARNED PUBLISHING Open Access

ORIGINAL ARTICLE OPEN ACCESS

Moving Open Repositories out of the Blind Spot of Initiatives to Correct the Scholarly Record

Frédérique Bordignon[1,2]

[1]Ecole nationale des ponts et chaussées, Institut Polytechnique de Paris, Marne-la-Vallée, France | [2]LISIS, INRAE, Université Gustave Eiffel, CNRS, Marne-la-Vallée, France

Correspondence: Frédérique Bordignon (frederique.bordignon@enpc.fr)

Received: 22 July 2024 | **Revised:** 29 October 2024 | **Accepted:** 28 November 2024

Funding: The NanoBubbles project has received Synergy grant funding from the European Research Council (ERC), within the European Union's Horizon 2020 program, grant agreement no. 951393.

Keywords: correction processes | open access | open repositories | scholarly communication

ABSTRACT

Open repositories were created to enhance access and visibility of scholarly publications, driven by open science ideals emphasising transparency and accessibility. However, they lack mechanisms to update the status of corrected or retracted publications, posing a threat to the integrity of the scholarly record. To explore the scope of the problem, a manually verified corpus was examined: we extracted all the entries in the Crossref × Retraction Watch database for which the publication date of the corrected or retracted document ranged from 2013 to 2023. This corresponded to 24,430 entries with a DOI, which we use to query Unpaywall and identify their possible indexing in HAL, an open repository (second largest institutional repository worldwide). In most cases (91%), HAL does not mention corrections. While the study needs broader scope, it highlights the necessity of improving the role of open repositories in correction processes with better curation practices. We discuss how harvesting operations and the interoperability of platforms can maintain the integrity of the entire scholarly record. Not only will the open repositories avoid damaging its reliability through ambiguous reporting, but on the contrary, they will also strengthen it.

1 | Introduction

When errors or misconduct are discovered in published research, journals issue editorial notices to correct the scholarly record. These notices, which include expressions of concern, corrections, additions, errata, corrigenda, withdrawals, and retractions, serve as a crucial mechanism for maintaining the integrity of the scientific literature. Retractions, in particular, are considered the most severe measure, often referred to as the 'nuclear option' for editors and publishers (Marcus and Oransky 2017). Editorial notices acknowledge and correct honest mistakes as well as instances of fraud and misconduct. As Dougherty (2019) observes, they represent a 'disruptive intervention upon the scholarly record', deviating from the normal trajectory of scientific advancement.

Not only do they correct specific errors, but their publication serves broader purposes, for example, they warn readers of issues with prior findings that may impact the interpretation of future studies; they also promote transparency around corrections to uphold the integrity and reliability of the scholarly record. It is very difficult if not 'an impossible task' (Lavoie et al. 2014) to delineate this scholarly record and identify its boundaries as it has undergone profound changes over time, starting with its extension from paper collections, which libraries are responsible for in physical spaces, to their electronic counterpart, in an infrastructure that is even more distributed than the global library network. Its size has increased, along with the diversity of the documents it contains, with the inclusion of ancillary products such as data, computer code, research notebooks and conference presentation materials. Following the Online Computer Library Center

Summary

- The open access movement has focused on disseminating knowledge and facilitating long-term archiving, without anticipating the need to later update the status of documents in repositories.
- For 91% of retracted/corrected publications deposited or indexed in HAL (second largest institutional repository worldwide), there is no statement of the new status of the publication.
- Open repositories should archive editorial notices to compensate for the frequent shortcomings of publishers in providing effectively such information.

(Lavoie et al. 2014), we adopt a broad definition of the scholarly record, including open repositories and the objects they contain.

In this article, we propose to explore the role of open repositories in a time when the scholarly record reliability is challenged by more retractions (Van Noorden 2011).

We will begin by conducting a literature review to explore the intricate nature of the scholarly record, and to highlight the ambiguous role of open repositories within it when corrections need to be made. Our review will show that open repositories have remained in the blind spot of initiatives seeking to enhance the reliability of scientific literature and the way scientific outputs are displayed online. Next, through a quantitative analysis of a corpus of retracted or corrected publications hosted in the HAL open repository, we aim to better understand the scope and magnitude of the issue. Finally, we will discuss the potential role that open repositories could play in contributing to correction of science and the integrity of the scholarly record.

2 | Background

2.1 | A Diverse and Distributed Scholarly Record

Roosendaal and Geurts (1997) identify four functions in scholarly communication: (1) registration, which establishes intellectual priority and assigns responsibility for the work produced, (2) awareness, which ensures dissemination, (3) archiving, which guarantees preservation, and (4) certification, which is intended to ensure quality or validity via control mechanisms, typically peer review.

The registration and archiving functions go beyond the performative act of publication by a publisher, and 'fix' (Cramer et al. 2023; Lavoie et al. 2014) any type of scientific production in the record. This includes material that has not necessarily transitioned through the peer-review process, such as preprints or datasets, or, on the contrary, content that results from the post-publication deconstruction of articles into 'fragments' (Boukacem-Zeghmouri 2021), such as the tables, images or maps from which they were drawn. Making them available to the community no longer depends solely on the work of the staff responsible for collecting, curating and disseminating the work, but is increasingly based on a network of online platforms and unique identifiers assigned to all types of objects, starting with the DOI. Open repositories and the objects they contain are therefore part of the scholarly record.

A DOI, or Digital Object Identifier, is a unique alphanumeric string assigned to a digital object, such as a journal article, dataset, or other scholarly content. The DOI provides a persistent and reliable link to the object, ensuring that it can always be found on the internet even if the location (URL) of the content changes. A DOI is a unique identifier assigned to a scientific production, but it is far from being its unique identifier: indeed, each platform on which it circulates (or its metadata circulate) assigns it an additional identifier in turn. These identifiers are all entry points into the record and when a publisher issues a correction or retraction notice, they play a key role in the dissemination of the information, since the content of the document remains intact, but its description (i.e., metadata conveyed by the DOI) changes to indicate the new status.

In a study conducted on a chemistry journal, Bordignon (2023) has highlighted a situation when two contradictory scenarios unfold simultaneously: on the publisher's website, a retraction may be poorly reported, while at the same time, the very same retraction is accurately reported on other platforms, such as PubMed (a bibliographic database in biology and medicine), PubPeer (a post-publication commentary platform) or Scite (a tool for detecting the polarity of citations, whether supporting, mentioning or contrasting). Numerous platforms, both open and commercial, draw on the record, or add to it, or both at the same time, and cross-reference their metadata to enrich them and inform users of the status of a document, even without the knowledge of the original producer: this is how the scholarly record can be maintained up to date and reliable. Institutional or private parties are in charge of this maintenance, for example: libraries with their document portals, academic social networks that make documents available (ResearchGate, Academia, Reddit, etc.), public or private publishers, and variously funded coalitions that develop repositories (Arxiv, COS, Zenodo, Dryad, Mendeley, etc.). Additionally, through acquisitions or mergers, content passes from hand to hand resulting in a shift in responsibility for its maintenance (Dowding 2016; NISO 2008).

As Van De Sompel and Treloar (2014) observed, all these platforms, be they private or backed by public institutions, excel at meeting the registration and awareness functions, but they do not all offer the same guarantees in terms of archiving (and consequently in terms of compliancy with open science standards), nor in terms of certification.

Ultimately, there is no technical or theoretical barrier that can exclude from the record the content that does not fully meet certain criteria. On the contrary, the ecosystem of scholarly communication has been designed to maximise gateways and connections between documents and between their metadata, regardless of their origin (public or private), their status (published or unpublished), or their type and format. In an idealised version of open science (to which, for example, Thibault et al. (2023) aspire in their call for an Open Science 2.0), this modular and dynamic archive would include all forms of discussions, including corrections at all stages of research, and

encourage researchers to give more credit to recording successive versions of their work (*record of versions* (Bosman 2016)), rather than a static official publication (*version of record* (NISO 2008)).

Consequently, for this heterogeneous archive, with its fuzzy boundaries and dematerialized content distributed across a large number of platforms, including open repositories, maintenance through correction operations remains a challenge, even though it is essential to the 'well-being of science' (Zuckerman 2020).

2.2 | Open Repositories to Maximise Access to the Literature

Open repositories have played a pivotal role in advancing open access to scholarly research. As defined by Pinfield (2009), a repository is a 'set of systems and services which facilitates the ingest, storage, management, retrieval, display, and reuse of digital objects'. Open repositories aim then to maximise the accessibility and impact of knowledge by providing open and sustained access to scholarly works.

Historically, early open archives emerged as socio-technical tools that improved and accelerated the circulation of preprints within specialist research communities, such as high-energy physicists on arXiv and economists on RePEc in the early 1990s (Mounier 2010; Pinfield et al. 2014). With the advent of the Internet, these repositories rapidly came to support the broader open access movement, translating its principles and philosophy into technical implementation as the pioneers of the movement recognised new digital communication technologies as a means to advance their ideological goals (Cohoon and Howison 2021; Moore 2017).

Repositories are designed to enable 'green' open access, that is where authors self-archive their own work (Gadd and Troll Covey 2019). Open repositories are considered a solution to two interrelated problems in scholarly communication: access and impact (Harnad et al. 2008). The access problem arose from escalating journal subscription costs (*serials crisis*), leading to library cancellations and scholars being unable to access the literature they needed. This, in turn, resulted in the impact problem, where scholars' work remained unread and uncited, reducing their scholarly impact.

Then, the open science movement both built upon and broadened the scope of open access. While comprising a diversity of aims, open science is structured around voluntary principles of enhanced transparency and better scientific rigour (Lyon 2016; Thibault et al. 2023). These include making not only publications openly accessible, but also calling for research data, computer code, protocols, methods, and laboratory notebooks to be shared.

At the same time, in the early 2000s, voices were raised to sound the alarm about the reproducibility crisis (a methodological crisis arising from the observation that published results are too often difficult, if not impossible, to reproduce) and to point out cases of scientific misconduct (Artino, Driessen, and Maggio 2019; Biagioli and Lippman 2020; Crocker and Cooper 2011; Fanelli 2009; Ioannidis 2005; Prinz, Schlange, and Asadullah 2011), resulting in growing numbers of corrections and even retractions (Else 2024; Fang and Casadevall 2011; Steen, Casadevall, and Fang 2013; Van Noorden 2011).

However, while the number of repositories continues to rise (Bashir et al. 2022; Pinfield et al. 2014), there are no plans to manage these retractions and pass on the corrections issued by publishers. Research integrity experts and advocates of a reform of research practices are calling for better quality control of the scientific literature as it is produced, as well as improved reporting of anomalies identified post-publication (Anirban 2023; Cabanac 2022; Larousserie, Cabut, and Morin 2023): they are calling for scientific literature to be 'decontaminated' or 'cleaned up', and are very keen on the reliability of metadata and the reporting of corrections. They urge publishers to quickly correct and retract publications for which they identify problems. However, as far as we know, the institutions in charge of open repositories are not systematically asked to contribute to the correction effort.

To summarise, this literature review highlights the tension between two approaches: on one hand, proponents of open access focus on improving access to scientific publications through open repositories but often overlook the need for corrections; on the other hand, advocates for reforming research practices prioritise the reliability of the scientific record, yet tend to overlook the role of open repositories, primarily targeting publishers instead.

3 | Data and Methods

In Bordignon's study (2023), out of 1068 publications retracted from the same journal, there were only five reports of the retraction in the repositories hosting a full version. To go beyond the scope of a single journal, we took as our starting point the Retraction Watch database, whose data was recently opened thanks to a partnership with Crossref (2023), and which allows free querying via a DOI. Retraction Watch is an independent, non-profit organisation that monitors and reports on retractions of scientific papers. Founded in 2010, Retraction Watch aims to increase transparency in the scientific process by highlighting instances of misconduct, errors, or other issues leading to the retraction of research articles. Crossref is a non-profit organisation that provides a comprehensive citation-linking service for scholarly content. It is best known for assigning DOIs to research outputs, which ensure persistent and reliable links to digital objects. The data collection method is presented in Figure 1 and detailed below.

On 9 October 2023, we extracted all the entries in the Crossref × Retraction Watch database for which the publication date of the corrected or retracted document ranged from 2013 to 2023. This corresponded to 27,223 entries, including 24,430 with a DOI, which we used to query Unpaywall (a service for tracking open access status) and identify their possible indexing in an open repository. We found 7560 documents with at least one repository version, distributed across 369 distinct repositories identified by their URL domain names. The top 3 repositories

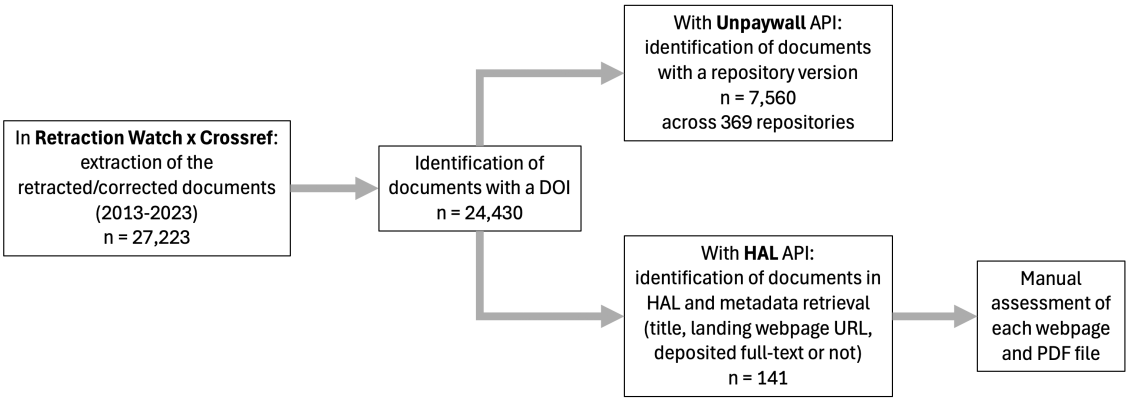

FIGURE 1 | Data collection method.

are PubMed ($n = 6728$), Europe PMC ($n = 3951$) and Semantic Scholar ($n = 1313$). These figures illustrated the potential scale of the issue if these documents are not properly flagged as corrected or retracted; but they were too numerous for us to carry out a manual check.

To demonstrate how the problem manifests itself and assess its scope, we have chosen to examine a corpus that allowed for manual verification. While a broader study would provide a more comprehensive assessment of the overall situation, such an expansion currently faces significant challenges. The heterogeneity of data across various online repositories makes automated collection and assessment impossible at present. It would require the development of sophisticated tools capable of navigating and interpreting the diverse data structures found across online repositories. Until such tools are available, a small but manually verified corpus offers valuable insights into the nature and scope of the problem. So, we chose the fourth repository in the list, HAL. HAL is an institutional repository managed primarily by CNRS, the French National Centre for Scientific Research, with significant involvement from other major French research institutions, including universities. It relies on a robust community of moderators and curators, mainly repository managers and academic librarians (casuHAL 2024). Importantly, HAL does not employ automatic harvesting operations from external sources. As a multidisciplinary institutional repository, HAL stands out as the second largest worldwide (Ranking Web of Repositories 2024). The only larger repository, NASA's Astrophysics Data System, focuses on a narrower range of disciplines. HAL's broad scope and substantial size made it an ideal subject for our study.

Instead of relying solely on Unpaywall results, we queried HAL (via its API) for the list of 24,430 retracted/corrected documents with a DOI, to ensure we didn't miss any that Unpaywall might not have indexed. We identified 141 documents in HAL, and retrieved the title, the landing webpage URL, and whether there is a deposited full-text or not (see Supporting Information for full list). We then manually checked whether the reader is alerted to their correction or retraction when consulting them on the platform. In concrete terms, this process involved two steps: (1) accessing the metadata of the document on the landing webpage; this page was examined for any mentions of a correction, (2) opening the PDF file, if available, and searching within the full text for marks indicating a correction or retraction. In 91% of cases, there was no mention of any correction in HAL (Table 1) with no significant difference depending on whether the full-text was deposited or if there was only descriptive metadata (Table 1).

TABLE 1 | Statement availability, out of 141 retracted/corrected publications in HAL, retrieved from Retraction Watch × Crossref database (2013–2023).

Statement availability	#	%
No correction statement	128	91%
Correction statement in the title	11	8%
Correction statement elsewhere	2	1%
All	141	100%

Differentiating by type of correction, it was for retractions that there were the most notifications (Table 2), notably through a mention in the title (13% of cases, which is still very low). As this procedure is not formalised, there are many variants such as the words *retracted* or *rétracté* (the French equivalent), inserted at the beginning or end of the title. Other strategies also co-exist in the database such as adding a watermark or mention on the archived version of the document.

Finally, we did not find any link to an editorial notice of correction, whether to a potentially archived version in HAL or to the publisher's notice. Our data are limited insofar as it is complicated to fully automate the process over a wider scope that would take into account differences in languages or the difficulty of detecting a watermark in a PDF file.

4 | Discussion

4.1 | Open Repositories in a One-Way Openness Movement

The examination of HAL confirms the tendency revealed by the literature review: whereas open repositories have solved the problems of access and impact, they have also aggravated the

TABLE 2 | Statement availability and type of correction, out of 141 retracted/corrected publications in HAL, retrieved from Retraction Watch × Crossref database (2013–2023)—The total is not the exact sum of the column as a document can be first corrected, then retracted.

Type of correction	Statement	#	%
Retraction	No statement	74	85%
	Statement in the title	11	13%
	Statement elsewhere	2	2%
	Subtotal	*87*	*100%*
Other type of correction	No statement	62	98%
	Statement in the title	1	2%
	Correction statement elsewhere	0	0%
	Subtotal	*63*	*100%*
All		141	100%

issue of the uncontrolled widespread dissemination of errors identified in the literature because they do not relay publishers' alerts, and do not effectively integrate correction or retraction notices that are published, whether by archiving them internally or by establishing the bidirectional link between the notice and the document as recommended by the Committee on Publication Ethics (Barbour et al. 2009) and the National Information Standards Organisation (NISO 2008). Originally created to facilitate the circulation of preprints and then used to support open access to publications, open access repositories were not designed with a view to managing post-publication corrections, and they have remained (for now) in the blind spot of initiatives promoting research integrity whose concerns go hand in hand with those about the integrity of the scholarly record (Daling 2023). Therefore, the open access movement can be described as unidirectional as it has focused solely on disseminating knowledge and facilitating long-term archiving, without anticipating the need to 'reverse course', that is, the need to later update the status of the documents indexed in repositories. The work of the Confederation of Open Access Repositories is very revealing in this regard: their technical recommendations for 'next generation repositories' (Boliini et al. 2017) provide nothing specific for addressing corrections and retractions.

Of the four functions of scholarly communication already mentioned above, it is the certification function that is lacking in open repositories, not because they do not contain certified objects, but because they are not designed to fulfil this function when a document is deposited there. Depending on the platform, certification may be granted *after* the deposit, resulting from the combination of micro-operations (likes, ratings, recommendations, 're-shares', or comments). This accumulation can contribute to the certification or not (Van De Sompel and Treloar 2014); but although it is carried out by peers, it does not reach the level of certification traditionally attributed to the peer-review of a scientific journal.

Function \ Venue	Publisher	vs	Open repository
Registration	++	=	++
Awareness	+/-	<	++
Archiving	+/-	<	++
Certification	++	>	-

FIGURE 2 | Functions of scholarly communication: Comparison between publishers' and open repositories' capacities (++ means optimal capacities for the function under scrutiny in both type of venues; – means no capacity at all; +/– means the capacities range from good to bad depending on the venue).

Figure 2 presents a comparative analysis of the capacities for registration, awareness, archiving, and certification across publisher venues and open repositories.

Regarding the registration function, the capacities are similar and optimal (++) in the performative act of publication on a journal platform, or in the claim of precedence enabled by a deposit in a repository. In terms of awareness and archiving, repositories have an optimal capacity by nature, whereas this can vary among publishers (+/–), whether in their degree of openness (from subscription-based to open access model) and the care they take in long-term archiving (see, e.g., this study (Laakso, Matthias, and Jahn 2021) which shows that even OA journals can disappear).

Open repositories are therefore not equipped (–) to carry out the 'decertification' constituted by a correction or retraction procedure. This explains why advocates of self-archiving and open-access repositories managers, key players in the opening of science, did not anticipate that the preserved output could be updated through correction operations. No specific metadata or unified procedure has been provided to flag corrections within them, or to contribute to better signalling in the record through cross-referencing of metadata.

4.2 | How Open Repositories Can Contribute to the Scholarly Record Reliability

To determine what role open repositories can play in ensuring the reliability of the scholarly record, we must first ask whether corrections should necessarily and systematically be reported in those repositories. This is a difficult question stemming from the fact that open repositories inherently destabilise the scholarly record; indeed, they are not a perfect reflection of

what has been officially published: they can archive a copy of the published official version (version of record) but also an 'unofficial' version from the author, either before acceptance by the journal (preprint), or after acceptance (postprint). The reason that led to a correction or retraction may not exist in the version deposited in an open archive, for example an error may have been introduced in a later version than the initial shared preprint. While preprints and published versions often differ only slightly (Klein et al. 2019), repository managers cannot be expected to compare documents and decide on correction mentions. These managers, academic librarians, and metadata curators are not peers invested with the authority to de-certify research; certification and decertification remain the prerogatives of publishers.

The primary need is to improve the status signalling of each archived version. Readers should know immediately whether they are viewing content identical to the published—and possibly retracted—version. This is particularly crucial because the same core metadata is associated with all versions (preprint, postprint, publisher version) and conveyed by the DOI.

Repository managers could play a key role in disseminating and archiving correction information. Open repositories distinguish themselves from journals through superior dissemination and archiving capabilities. These fundamental functions position them to contribute significantly to maintaining the integrity of the scholarly record.

Since publishers do not always report corrections in an exemplary manner, open repositories can address this gap. They can archive editorial notices, update publication notices with dedicated metadata, and provide visible, bidirectional technical links between corrections and original publications. For DOI-based publications, this process could be automated and facilitated by the Crossref × Retraction Watch partnership. It could also leverage information from other platforms such as PubMed, PubPeer or Scite (mentioned above), which employ different sources and retraction detection methods.

It should be noted that for retractions only, following the mechanism recommended to publishers, the file does not have to be removed (NISO 2008). However, in exceptional and extremely limited cases, it may be necessary to remove it, if it is clearly defamatory, violates privacy, is the subject of a court decision or may present a serious risk to public health. The repository holding a copy of the offending publication must also delete the file so that it is no longer accessible via any platform. Finally, PhD theses are often disseminated via the open repository of the country or institutions in which they were published. The case of theses must be treated differently from other types of publication insofar as the document disseminated is associated with the awarding of a diploma. It is therefore in conjunction with the awarding body that the alert should be established, and the manuscript may be withdrawn altogether.

Through harvesting operations and the interoperability of platforms, the integrity of the entire scholarly record can be maintained, and not only will the open repositories not have damaged its reliability through ambiguous reporting, but on the contrary, they will have strengthened it. Lastly, they will remain in line with their long-standing fight against the supremacy of publishers, who are not always very clear about reporting corrections, especially when these are the result of breaches of integrity (Bordignon 2023; Schneider et al. 2020).

5 | Conclusion and Perspectives

Open repositories have historically evolved with the primary goal of enhancing access to scholarly publications and increasing their visibility within the academic community. This development has been driven by the ideals of open science, which emphasise transparency, accessibility, and the widespread dissemination of knowledge. However, this forward momentum has not accounted for the need to update the status of publications that have been retracted due to errors, ethical breaches, or other issues undermining their integrity. Consequently, open repositories find themselves in a one-way trajectory of openness, lacking mechanisms to effectively reverse course and slow access to compromised research. This absence of a 'reverse gear' poses challenges for maintaining the integrity of the scholarly record and ensuring that the academic community does not continue to disseminate and rely on invalidated findings. To contribute to the integrity of the scholarly record (and not weaken it), open repositories must fully deploy their archiving and dissemination functions: they must include correction and retraction notices in their objects and apply the same procedures that publishers are supposed to implement according to COPE and NISO recommendations, namely the clear display of a document's new status and a permanent bidirectional link between the corrected document and the notice signalling it. Similarly, by leveraging platform interoperability, mainly implemented through DOIs, correction information should also be propagated to associated objects, for example to datasets whose reliability may have been called into question.

Keeping in mind that open science practices do not guarantee protection against breaches of research integrity, it is also necessary to anticipate the need to signal corrections in virtuous and completely open models of publication such as overlay journals (whose publishing model is based on manuscripts deposited in open access repositories), or at the level of preprint certification infrastructures systems (like Peer Community In, preLights, PREreview).

Even though HAL is the world's second largest institutional repository, our study would need to be expanded to other repositories; but we hope it can be considered a first step toward improving the role of open repositories in correcting science processes, a necessary awareness when one sees that no improvement is envisioned in the very latest reports from relevant actors, such as the report (Stern et al. 2023) about 'responsible publishing' issued by cOAlition S (a coalition of funding agencies encouraging green open access) or the report about the 'Current State and Future Directions for Open Repositories in Europe' published by a consortium of European actors committed to open science development (Boliini et al. 2017).

Author Contributions

F.B. conceived the project, developed the methodology, and wrote the article.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

The data that supports the findings of this study are available in the Supporting Information of this article.

References

Anirban, A. 2023. "How Can We Make Science More Rigorous?" *Nature Reviews Physics* 5, no. 9: 498–499. https://doi.org/10.1038/s42254-023-00631-x.

Artino, A. R., E. W. Driessen, and L. A. Maggio. 2019. "Ethical Shades of Gray: International Frequency of Scientific Misconduct and Questionable Research Practices in Health Professions Education." *Academic Medicine: Journal of the Association of American Medical Colleges* 94, no. 1: 76–84. https://doi.org/10.1097/ACM.0000000000002412.

Barbour, V., S. Kleinert, E. Wager, and S. Yentis. 2009. "Guidelines for Retracting Articles. Committee on Publication Ethics." https://doi.org/10.24318/cope.2019.1.4.

Bashir, S., S. Gul, S. Bashir, N. T. Nisa, and S. A. Ganaie. 2022. "Evolution of Institutional Repositories: Managing Institutional Research Output to Remove the Gap of Academic Elitism." *Journal of Librarianship and Information Science* 54, no. 3: 518–531. https://doi.org/10.1177/09610006211009592.

Biagioli, M., and A. Lippman, eds. 2020. *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. Cambridge, MA: MIT Press.

Boliini, A., P. Knoth, P. Perakakis, et al. 2017. "Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group (2017)." COAR. https://doi.org/10.5281/zenodo.8077381.

Bordignon, F. 2023. "What Does It Mean to Correct the Scientific Record? A Case Study of the JACS (2000–2023)." In *27th International Conference on Science, Technology and Innovation Indicators. STI 2023, Leiden, The Netherlands*. https://doi.org/10.55835/643e9f61c0a5d2717e40e59c.

Bosman, J. 2016. "@TAC_NISO Indeed, We Need Full Adoption of the Idea of a *Record of Versions*, Through All Workflow Phases and Crossplatform #Stminnovations. X (Formerly Twitter)." https://twitter.com/jeroenbosman/status/806467999620362240.

Boukacem-Zeghmouri, C. 2021. "La fin de la publication scientifique? Une analyse entre légitimité, prédation et automatisation. Histoire de la recherche contemporaine. La revue du Comité pour l'histoire du CNRS, Tome X-no 2, Article Tome X-no 2." https://doi.org/10.4000/hrc.6184.

Cabanac, G. 2022. "Decontamination of the Scientific Literature." arXiv. Last modified October 28. https://doi.org/10.48550/arXiv.2210.15912.

casuHAL. 2024. "Le club utilisateur de l'archive ouverte HAL." Accessed January 25, 2025. https://www.casuhal.org/.

Cohoon, J., and J. Howison. 2021. "Norms and Open Systems in Open Science." *Information & Culture* 56, no. 2: 115–137. https://doi.org/10.7560/IC56201.

Cramer, T., C. German, N. Jefferies, and A. Wise. 2023. "A Perpetual Motion Machine: The Preserved Digital Scholarly Record." *Learned Publishing* 36, no. 2: 312–318. https://doi.org/10.1002/leap.1494.

Crocker, J., and M. L. Cooper. 2011. "Addressing Scientific Fraud." *Science* 334, no. 6060: 1182. https://doi.org/10.1126/science.1216775.

Crossref. 2023. "Crossref Acquires Retraction Watch Data and Opens It for the Scientific Community. Crossref." https://doi.org/10.13003/c23rw1d9.

Daling, D. 2023. "'On the Ruins of Seriality': The Scientific Journal and the Nature of the Scientific Life." *Endeavour* 47, no. 4: 100885. https://doi.org/10.1016/j.endeavour.2023.100885.

Dougherty, M. V. 2019. "Correcting the Scholarly Record for Research Integrity." *Monash Bioethics Review* 37, no. 1–2: 76–78. https://doi.org/10.1007/s40592-019-00093-x.

Dowding, H. 2016. "Open Access, Digital Preservation, and the Changing Scholarly Record." In *Open Access and the Future of Academic Libraries*, edited by K. Smith and K. Dickson, 203. Lanham, MD: Rowman & Littlefield.

Else, H. 2024. "Biomedical Paper Retractions Have Quadrupled in 20 Years—Why?" *Nature* 630: 280–281. https://doi.org/10.1038/d41586-024-01609-0.

Fanelli, D. 2009. "How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data." *PLoS One* 4, no. 5: e5738. https://doi.org/10.1371/journal.pone.0005738.

Fang, F. C., and A. Casadevall. 2011. "Retracted Science and the Retraction Index." *Infection and Immunity* 79, no. 10: 3855–3859. https://doi.org/10.1128/IAI.05661-11.

Gadd, E., and D. Troll Covey. 2019. "What Does 'Green' Open Access Mean? Tracking Twelve Years of Changes to Journal Publisher Self-Archiving Policies." *Journal of Librarianship and Information Science* 51, no. 1: 106–122. https://doi.org/10.1177/0961000616657406.

Harnad, S., T. Brody, F. Vallières, et al. 2008. "The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update." *Serials Review* 34, no. 1: 36–40. https://doi.org/10.1080/00987913.2008.10765150.

Ioannidis, J. P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2, no. 8: e124. https://doi.org/10.1371/journal.pmed.0020124.

Klein, M., P. Broadwell, S. E. Farb, and T. Grappone. 2019. "Comparing Published Scientific Journal Articles to Their Pre-Print Versions." *International Journal on Digital Libraries* 20, no. 4: 335–350. https://doi.org/10.1007/s00799-018-0234-1 .

Laakso, M., L. Matthias, and N. Jahn. 2021. "Open Is Not Forever: A Study of Vanished Open Access Journals." *Journal of the Association for Information Science and Technology* 72, no. 9: 1099–1112. https://doi.org/10.1002/asi.24460.

Larousserie, D., S. Cabut, and H. Morin. 2023. "Chasseurs de fraude—Série de portraits." https://www.lemonde.fr/chasseurs-de-fraude/.

Lavoie, B., E. Childress, R. Erway, et al. 2014. "The Evolving Scholarly Record." Dublin, Ohio: OCLC Research. https://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-evolving-scholarly-record-2014.pdf.

Lyon, L. 2016. "Transparency: The Emerging Third Dimension of Open Science and Open Data." *LIBER Quarterly: The Journal of the Association of European Research Libraries* 25, no. 4: 4. https://doi.org/10.18352/lq.10113.

Marcus, A., and I. Oransky. 2017. "Is There a Retraction Problem? And, If So, What Can We Do About It?" In *The Oxford Handbook of the Science of Science Communication*, edited by K. H. Jamieson, D. M. Kahan, and D. A. Scheufele. Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190497620.013.13.

Moore, S. A. 2017. "A Genealogy of Open Access: Negotiations Between Openness and Access to Research." *Revue française Des Sciences de l'information et de la Communication* 11. https://doi.org/10.4000/rfsic.3220.

Mounier, P. 2010. "Le libre accès: Entre idéal et nécessité." *Hermès, La Revue* 57, no. 2: 23–30. https://doi.org/10.4267/2042/38634.

NISO. 2008. "NISO RP-8-2008, Journal Article Versions (JAV): Recommendations of the NISO/ALPSP JAV Technical Working Group." NISO. https://doi.org/10.3789/niso-rp-8-2008.

Pinfield, S. 2009. "Journals and Repositories: An Evolving Relationship?" *Learned Publishing* 22, no. 3: 165–175. https://doi.org/10.1087/2009302.

Pinfield, S., J. Salter, P. A. Bath, et al. 2014. "Open-Access Repositories Worldwide, 2005–2012: Past Growth, Current Characteristics, and Future Possibilities." *Journal of the Association for Information Science and Technology* 65, no. 12: 2404–2421. https://doi.org/10.1002/asi.23131.

Prinz, F., T. Schlange, and K. Asadullah. 2011. "Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets? Nature Reviews Drug Discovery." *Art* 10, no. 9: 712. https://doi.org/10.1038/nrd3439-c1.

Ranking Web of Repositories. 2024. "Transparent Ranking: Institutional Repositories by Google Scholar (March 2024)." Ranking Web of Repositories. https://repositories.webometrics.info/en/institutional.

Roosendaal, H. E., and P. A. T. M. Geurts. 1997. "Forces and Functions in Scientific Communication: An Analysis of Their Interplay," in *Conference on 'Co-Operative Research in Information Systems in Physics,' Oldenburg, Germany.* https://research.utwente.nl/en/publications/forces-and-functions-in-scientific-communication-an-analysis-of-t.

Schneider, J., D. Ye, A. M. Hill, and A. S. Whitehorn. 2020. "Continued Post-Retraction Citation of a Fraudulent Clinical Trial Report, 11 Years After It Was Retracted for Falsifying Data." *Scientometrics* 125, no. 3: 2877–2913. https://doi.org/10.1007/s11192-020-03631-1.

Steen, R. G., A. Casadevall, and F. C. Fang. 2013. "Why Has the Number of Scientific Retractions Increased?" *PLoS One* 8, no. 7: e68397. https://doi.org/10.1371/journal.pone.0068397.

Stern, B., Z. Ancion, A. Björke, et al. 2023. "Towards an Open, Scholar-Led Communication Ecosystem: A New Proposal From cOAlition S." https://doi.org/10.5281/zenodo.8398480.

Thibault, R. T., O. B. Amaral, F. Argolo, A. E. Bandrowski, R. D. Alexandra, and N. Drude I. 2023. "Open Science 2.0: Towards a Truly Collaborative Research Ecosystem." *PLoS Biology* 21, no. 10: e3002362. https://doi.org/10.1371/journal.pbio.3002362.

Van De Sompel, H., and A. Treloar. 2014. "A Perspective on Archiving the Scholarly Web," in *iPres 2014 Conference*, 197.

Van Noorden, R. 2011. "Science Publishing: The Trouble With Retractions." *Nature* 478, no. 7367: 26–28. https://doi.org/10.1038/478026a.

Zuckerman, H. 2020. "Is 'the Time Ripe' for Quantitative Research on Misconduct in Science?" *Quantitative Science Studies* 1, no. 3: 945–958. https://doi.org/10.1162/qss_a_00065.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.