

Retraction without Traction: Towards Handling Post-publication Invalidation in Open Science Infrastructure

Christoph Stade^{*,†}, Dexter Williams^{*,†}

The Information School, University of Wisconsin-Madison, Madison, WI, USA

Abstract

The process of retraction can invalidate the results and claims of a scientific publication. Therefore, to maintain the integrity of past and future research, it is critical for human researchers and research agents to know when a work has been retracted. Yet scholarly knowledge graphs like OpenAIRE fail to consistently communicate retraction status. We argue for the inclusion of an explicit retraction indicator in OpenAIRE's schema by extending it with scholarly ontologies that represent retraction information; this retraction information is already available through a trusted aggregation source: Crossref. Further, we borrow the concept of forgetting from early knowledge representation and cognitive psychology literature to conceptualize retraction within scientific knowledge graphs. Finally, we argue that retracted publications propagate uncertainty through a citation network and propose Kerchova & Dooren's PageTrust algorithm to compute this effect.

Keywords

Semantic publishing, Open science, Retraction, Scholarly knowledge graphs

1. Introduction

AI-driven research assistants are becoming a primary interface to scholarly knowledge [1], retrieving, ranking, and summarizing publications with increasing autonomy [2]. Their reliability, however, is critically bounded by the quality of scholarly data they consume. A recent evaluation of nine popular generative artificial intelligence tools, including those specifically oriented at research, showed that no tool consistently recognized whether a paper had been retracted [3]. Since their evaluation set consisted of frontier models, unreliability is not due to capability of the language model but metadata quality and coverage.

In scientific research, retractions are an important post-publication mechanism for addressing a wide range of issues and errors in published output. While a paper retracted for missing ethics approval might still report valid findings, one retracted for fabricated data does not. Through citation, invalidated claims of the latter propagate into the citation network as second-generation errors [4]. Post-retraction citation can continue for years after a paper's initial retraction [5, 6] and can support the paper's claims with undue validity to the point of harming the public [7].

Research agents access the scholarly record primarily through aggregated metadata rather than publications themselves; several tools evaluated by Labenbacher et al. [3] operate on corpora derived from scholarly knowledge graphs (SKGs). *Consensus*, for instance, uses the Semantic Scholar corpus.¹ Scholarly APIs may also be directly exposed to research agents through interfaces, such as Model Context Protocol servers.

OpenAIRE, a project by the European Union toward a federated open science infrastructure [8, 9], presents an opportunity to serve as an authoritative, centralized registry for tracking retractions. Unlike other indexing services, such as Google Scholar, OpenAIRE explicitly commits itself to data quality

6th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment, Oct 2026, Bari, Italy

*Corresponding author.

[†]These authors contributed equally.

✉ cstade@wisc.edu (C. Stade); djwilliams22@wisc.edu (D. Williams)

ORCID 0009-0009-3973-5139 (C. Stade); 0009-0002-0147-6628 (D. Williams)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://help.consensus.app/en/articles/10055108-consensus-research-database>

in its acquisition policy.² At the time of writing, however, the OpenAIRE graph and its schema do not consider retractions. A retraction surfaces, at best, as a “RETRACTED:” prefix imported into a publication’s title upon re-indexing. This mechanism is subject to delays of up to one month,³ is not queryable as metadata, and lacks a link between the publication and its retraction notice. Critically, the missing information is already openly available through Crossref, one of OpenAIRE’s aggregation sources.⁴ Yet OpenAIRE’s current Crossref mapping drops all fields carrying retraction status. This type of flattening is not a unique fault of OpenAIRE, for instance, OpenAlex prunes the same Crossref update metadata into a single boolean flag, misclassifying papers as (non-) retracted [11]. Because SKGs such as OpenAIRE seek to become the structural backbone of scholarly communication, they have a critical need for data quality. Science needs trustworthy, machine-actionable representations of scholarly knowledge for research intelligence such as AI-driven research assistants; this includes information on retraction.

In this paper, we argue for explicit handling of retractions within scholarly knowledge graphs using OpenAIRE as a paradigmatic example. We conceptualize retractions as a form of *forgetting*. An initial retraction is an *ignorance*: a retracted publication must not be removed from the record but should lose its license to support further work. We argue that this ignorance can be incorporated into OpenAIRE at the ontology level by enriching OpenAIRE’s Crossref mapping to retain dropped UPDATE fields, making retraction a first-class, machine-actionable part of the graph. A second form of forgetting occurs as a consequence of retraction: papers that cite retracted works are *weakened*. We propose that the propagation of weakening can be implemented at the network level using Kerchov & Dooren’s *PageTrust* algorithm [12].

2. Background

2.1. Retraction in scholarly knowledge graph systems

Open scholarly knowledge graphs aim to centralize and make publicly available representations of scientific data as published artifacts in machine-actionable metadata formats. Examples are OpenAlex, which inherits the Microsoft Academic Knowledge Graph (discontinued in 2021) [13], and OpenAIRE, an initiative by the European Union [14].

Few popular representation models for structured scientific data have mechanisms to discard invalid information. Retracted nanopublications disappear from default search results [15].⁵ Similarly, Wikidata statements with deprecated rank leave the *truthy* WDT: namespace, which can be explicitly queried via SPARQL to omit invalidated statements.⁶ Both existing mechanisms are a naive filtering; i.e., the artifact leaves the standard search interface. In neither mechanism are publications that cite retracted publications affected and no consequences carry into the citation network.

OpenAIRE implicitly provides two ways to discover retractions. The first is manually by following one of the external version links. Another, more explicit way, is to discover a “RETRACTED:” title flag, appearing in OpenAIRE after the service re-indexed its sources.⁷ However, external links and title flags are not distinct markers of retraction status as they occur within existing metadata fields; therefore, external links and title flags are insufficient for signaling retraction status. Additionally, the relationship between the publication and its retraction notice is not established. Thus, retraction status may go unnoticed for both humans and AI agents. Additionally, re-indexing can have delays of up to one month.³ OpenAIRE makes pledges toward data quality in its content acquisition policy,⁸ which

²<https://www.openaire.eu/content-acquisition-policy>

³<https://www.openaire.eu/the-openaire-graph-why-continuous-validation-matters>

⁴Crossref acquired the Retraction Watch database in 2023 and, since 2025, exposes retraction information in its REST API through the UPDATE-TYPE and UPDATE-TO fields [10].

⁵<https://nanopub.readthedocs.io/en/latest/publishing/retraction.html>

⁶https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format

⁷For example, see <https://explore.openaire.eu/search/publication?pid=10.1101%5C%2F2020.03.16.20037135> or <https://explore.openaire.eu/search/publication?pid=10.1038%2Fnature00870>

⁸<https://www.openaire.eu/content-acquisition-policy>

becomes a danger to any researcher or AI agent that unsuspectingly consumes OpenAIRE records as a reliable resource. Challenges with communicating retractions extend beyond OpenAIRE; for example, retraction indexing databases have been found to disagree over the retracted status of individual papers [16]. Inconsistencies in the communication of retraction status led the *National Information Standards Organization* (NISO) to issue recommended practices for communicating retractions, removals, and expressions of concern [17].

The limited expressivity of OpenAIRE — skipping retraction status — can be argued for from an alignment perspective: The OpenAIRE data model⁹ is designed as the lowest common denominator across aggregation sources. However, strictly adhering to this minimal baseline excludes post-publication metadata critical to research integrity that OpenAIRE could readily integrate. Since January 2025 Crossref, one of OpenAIRE’s aggregation sources, provides retraction information via its REST API [10]. This information is recorded as the `UPDATE-TYPE` metadata field with the value `RETRACTION`. The target of a retraction notice (the retracted article) is then specified as `UPDATE-TO`. Who issued/recorded the retraction, the `SOURCE`, can be either `PUBLISHER` or `RETRACTION-WATCH`. Currently, several fields of Crossref’s JSON schema are dropped by OpenAIRE, including `UPDATE-TO`.¹⁰

2.2. Vocabulary to retract

At the level of data model, several ontologies of the *Semantic Publishing and Referencing Ontologies* (SPAR) family provide vocabulary to describe retractions [18]. The Citation Typing Ontology (CiTO) [19] defines the act of retracting as a relationship between a publication (retracted) and a retraction notice (retracting) with an inverse property pair: `CITO:RETRACTS` and `CITO:ISRETRACTEDBY`. FaBiO [19] adds document types with `FABIO:RETRACTION` for the retraction statement as an abstract work, and `FABIO:RETRACTIONNOTICE` for the published expression that realizes it. Further, `FABIO:CORRECTION` has the subclasses `FABIO:CORRIGENDUM` and `FABIO:ERRATUM` to separate author-side errors from the publisher [19]. The property `FABIO:HASRETRACTIONDATE` takes note when a retraction took effect, which allows for a split between pre-retraction and post-retraction in a publication’s life cycle. From the Publishing Status Ontology (PSO) ontology, the named individual `PSO:RETRACTED-FROM-PUBLICATION` describes the status of being retracted. This status can be attached to a publication with `PSO:HOLDSSTATUSINTIME` and time-indexed [20]. None of the SPAR ontologies model specific reasons for retraction.

OpenAIRE’s relationship semantics incorporate the DataCite Metadata Schema [21]. By aggregating over this schema OpenAIRE embeds all the citation ontologies mentioned in this section.^{11,12} The closest to a retraction in the DataCite schema is an additional `DATE` to the publication that is typed with `DATETYPE` as “Withdrawn”.

Finally, it should be noted that OpenAIRE does not carry forward the ontology definitions by specifying namespaces. Relationships are simply strings inspired by an underlying controlled vocabulary, but are not controlled themselves. This is a consequence of OpenAIRE stopping its Linked Open Data (LOD) service that exposed its graph as *Resource Description Framework* through a SPARQL endpoint.^{13,14}

3. Retractions as forgetting

To better illustrate the epistemic consequences of retraction, we borrow the notion of *forgetting* from early knowledge representation and cognitive psychology literature. Different forms of forgetting are based on different cognitive processes and can be characterized by their outcome [22]. We understand retraction as a process of belief revision and a direct and intentional contraction on claims made in the scientific ecosystem. However, retraction is not revocation; the claims of a retracted publication are

⁹<https://graph.openaire.eu/docs/>

¹⁰https://graph.openaire.eu/docs/graph-production-workflow/aggregation/non-compatible-sources/crossref_unpaywall

¹¹<https://graph.openaire.eu/docs/graph-production-workflow/aggregation/non-compatible-sources/datacite/>

¹²<https://graph.openaire.eu/docs/data-model/relationships/relationship-types>

¹³<https://www.openaire.eu/pausing-our-lod-services>

¹⁴<https://graph.openaire.eu/docs/downloads/related-datasets/#the-openaire-lod-dataset>

ignored rather than negated. A retracted publication should not be hidden, yet must become unavailable in some other way, namely citation. Thus, retraction is a process of *ignorance* [22].

Formulating retraction as a process of ignorance is informative for understanding how retracted publications propagate uncertainty in a network of scientific knowledge. The references of a publication can be seen as a part of the epistemic support for its claims. When a referenced publication is retracted the citing publication becomes ignorant of the referenced claims of the referenced publication. While a publication citing a retracted publications does not undergo a process of ignorance, it does undergo a process of *weakening*; this process of weakening propagates, as publications that reference weakened publications become weakened themselves. In all, according to our conceptualization of retraction, SKGs should both capture the initial retraction (ignorance) and the propagation of uncertainty (weakening).

Relation	Crossref metadata	Domain	Range
CITO:ISRETRACTEDBY	DOI	RESEARCHPRODUCT	FABIO:RETRACTIONNOTICE
FABIO:HASRETRACTION-DATE	UPDATE-TO[UPDATED]	RESEARCHPRODUCT	XSD:DATETIME
PSO:HOLDSTATUSINTIME	<i>computed</i>	RESEARCHPRODUCT	PSO:STATUSINTIME (PSO:RETRACTED-FROM-PUBLICATION)
HASRETRACTIONTYPE	UPDATE-TO[TYPE]	PSO:STATUSINTIME	Subset of Crossref enumeration type CM_UPDATE_TYPE
HASRETRACTIONREASON	<i>mapped from Retraction Watch taxonomy</i>	PSO:STATUSINTIME	Retraction Watch taxonomy
CITO:RETRACTS	UPDATE-TO[DOI]	FABIO:RETRACTIONNOTICE	RESEARCHPRODUCT
RDFS:SUBCLASSOF	–	FABIO:RETRACTIONNOTICE	RESEARCHPRODUCT
HASSOURCE	SOURCE	FABIO:RETRACTIONNOTICE	{ PUBLISHER, RETRACTION-WATCH }

Table 1

Proposed extension to the OpenAIRE aggregation schema, grouped by the object to attach to (i.e. subject position), either a retracted publication or retraction notice. Because of the LOD limitations mentioned in Section 2.2, the semantics are implicit only (but recoverable).

3.1. Ignorance in OpenAIRE

A first tier of forgetting can be achieved by making the relationship between a retracted paper and its retraction notice explicit, similarly to the nanopublication approach [15]. For this end, we extend OpenAIRE’s Crossref mapping used in aggregation.¹⁵ The annotation process first needs to filter for retraction notices of the appropriate type (\$TYPE).¹⁶ We specifically look for retraction types consistent with our definitions of ignorance and weakening, which are, RETRACTION, REMOVAL, WITHDRAWAL and PARTIAL-RETRACTION.¹⁷ Then, we collect the DOI in each retraction notice’s UPDATE-TO[DOI] field, which is pointing to the retracted research product. For the publication represented by this DOI, we then populate the relations listed in Table 1. The table lists our naive extension to the OpenAIRE graph data model with information that is readily available through its aggregation sources, grouped by the object to attach to (retracted publication, retraction notice). The prefix left of the colon denotes from

¹⁵https://graph.openaire.eu/docs/graph-production-workflow/aggregation/non-compatible-sources/crossref_unpaywall

¹⁶[https://api.crossref.org/v1/works?filter=update-type:\\$TYPE](https://api.crossref.org/v1/works?filter=update-type:$TYPE)

¹⁷Subset of this enumeration type: https://data.crossref.org/reports/help/schema_doc/5.5.0/index.html#update. In the future, this set may be enriched with more nuanced reasons of Retraction Watch taxonomy like *Paper Mill* from <https://retractionwatch.com/retraction-watch-database-user-guide/retraction-watch-database-user-guide-appendix-b-reasons/>.

which (scholarly) ontology we import the vocabulary. A missing prefix indicates when we define a new vocabulary element. The *Crossref metadata* column specifies which Crossref metadata field of a retraction notice populates the relation. Brackets indicate a subfield (i.e., UPDATE[TYPE]).

We acknowledge OpenAIRE’s decision to retire its LOD service by proposing a lightweight schema extension that is technically non-semantic yet grounded in the scholarly ontologies mentioned in Section 2.2. This way we make the vocabulary recoverable, i.e., liftable to RDF by third parties.

Listing 1 illustrates the extended schema for a concrete example research product in JSON-LD format. Removing the *@context* list transforms the example into plain JSON, which is the format that the OpenAIRE API serves. Our example moves a publication’s retraction notice away from its deduplicated representative entity; at the moment, for some retracted publications the retraction notice will be indexed as simply a version of the publication. This way the retraction notice will live as a distinct resource in the graph and have its retraction relationship be made explicit.

3.2. Weakening in OpenAIRE

According to our conceptualization of retraction as forgetting, publications that either cite retracted or weakened publications are *weakened*. In other words, a retracted publication propagates uncertainty throughout an SKG. OpenAIRE captures citation relationships between papers, allowing the construction of a directed citation network $\mathcal{G}(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ are research products, and $\mathcal{L}$ are asymmetric ISCITEDBY relationships. To represent a node’s, and thus a research product’s, importance within $\mathcal{G}$, any number of classic node centrality metrics can be used to calculate a rank value. A well-known node ranking algorithm, which OpenAIRE natively supports,¹⁸ is *PageRank* [23]. Rankings produced by PageRank can be interpreted as the frequency of random walkers who follow positive links visiting the node, with a chance for a random jump. PageRank is calculated iteratively: nodes gain rank based on their in-degree, and in turn confer rank to nodes they connect to. When PageRank is applied to a citation network $\mathcal{G}$ it can be viewed as a rank of a publication’s importance based on the number and quality of its ISCITEDBY relationships.

While PageRank can be applied to $\mathcal{G}$, consider the alternative citation network $\mathcal{G}'(\mathcal{N}, \mathcal{R}, \mathcal{L}^+, \mathcal{L}^-)$, where $\mathcal{N}$ are publications, $\mathcal{R} \subset \mathcal{N}$ are retracted publications, $\mathcal{L}^+$ are ISCITEDBY relationships with a source node in $\mathcal{N} \setminus \mathcal{R}$, and $\mathcal{L}^-$ are ISCITEDBY relationships with a source node in $\mathcal{R}$. A successful ranking algorithm applied to $\mathcal{G}'$ should drive down the rank of publications citing retracted publications; successfully capturing *weakening*, by conferring lower ranks to publications that rely on weakened publications.

Unfortunately, PageRank is not able to handle negative edge weights and either ignores them or views them positively. Therefore, we tentatively propose applying *PageTrust* [12], a modification to PageRank that accounts for negative edge weights, to $\mathcal{G}'$ in order to measure weakening in OpenAIRE. Rankings produced via PageTrust can be interpreted as the frequency of random walkers visiting the node who have not accumulated distrust towards the node; random walkers accumulate distrust towards a node when they follow a negatively weighted edge. Similar to PageRank, PageTrust computes rankings iteratively and propagates rankings throughout a network.

In Figure 1, we provide an example of PageTrust applied to a topical subgraph of OpenAIRE on COVID-19 research with 921,422 publications, 7,008,976 citations and 911 retracted papers.¹⁹ We define weakening as $w = 1 - \frac{\text{PageTrust}}{\text{PageRank}}$. We call a publication *weakened* when its PageTrust is below its PageRank ($w > 0$). This weakened status communicates that a publication is likely unreliable and should be cited with caution. Of the 921,422 nodes in the OpenAIRE subset that have an in-corpus reference (471,713), i.e. cite another publication in this subset, 34,342 (7.3%) are weakened, all within four citation-hops of a cited retracted publication.

In Figure 1(a), communities 4 and 18 are most weakened, with 17.1% and 16.2% of their members weakened. Both have about 7.5% of members citing a retracted publication directly. Community 4

¹⁸<https://bip.imsi.athenarc.gr/site/data>

¹⁹Code to reproduce the figure can be found at https://github.com/dexterwilliams96/pagetrust_sci_k

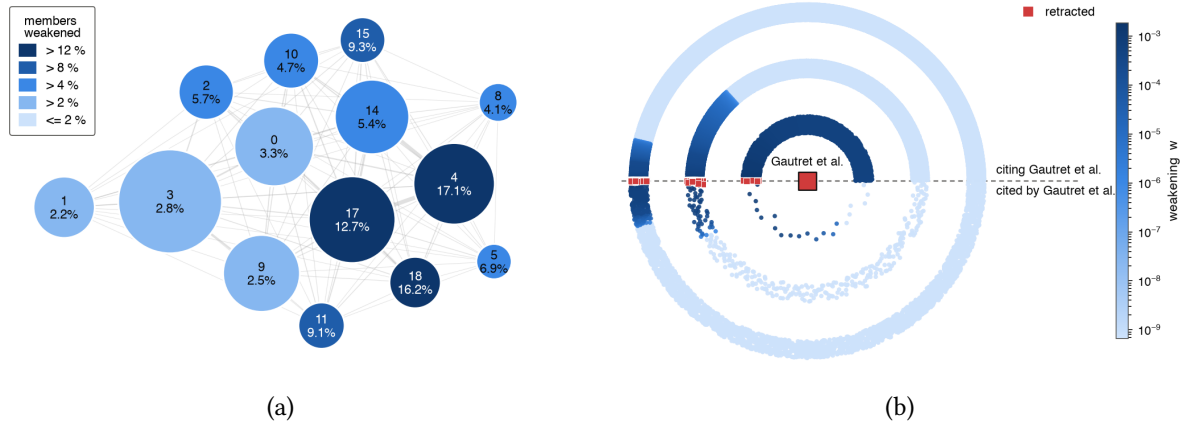


Figure 1: PageTrust weakening computed on publications of the COVID-19 OpenAIRE subset [25] that have in-corpus references (471, 713). (a) partitions every publication by Leiden modularity ($resolution = 1$) on the undirected direct citation graph. We only show the 14 citation communities with at least 1,000 members. Edges are citations between communities, with width proportional to the count. Each circle’s size corresponds to the number of community members and its shade of blue is relative to the percentage of non-retracted members that are weakened. (b) centers on the notorious and highly cited retracted paper by Gautret et al. [24]. Three rings of citing and cited papers (as dots) are placed around Gautret et al. [24], separated horizontally by citation direction (cited by: 3, 507, 54, 568 and 172, 468; citing: 21, 328 and 2, 859). A ring’s distance to the center point signifies the number of citation-hops. Each dot is colored relative to its weakening w . For readability, we place color distances on a logarithmic scale and cap the darkest shade at the 99.5th percentile of non-retracted papers. Within each half-ring, the dots are sorted by w . Red squares are retracted papers.

contains Gautret et al. [24], and 2,182 of its 4,100 direct citers of retracted work cite this paper (4.0% of the community, compared to 0.2% in community 18). Gautret et al. [24] substantially determines the weakening of community 4. However, the number of retracted publications in a community does not predict its weakening as community 0 contains 113 retracted publications with a weakening of 3.3% whereas community 18 has just 28 retractions yet a weakening of 16.2%. Overall, weakening is a thin signal, as a weakened publication loses just 10^{-4} of its rank. Our results suggest that the PageTrust algorithm distributes weakening non-arbitrarily. Weakening concentrates in communities that build most directly on retracted work. We therefore see PageTrust as a reasonable starting point for computing how the effects of retraction propagate through a citation network.

Our approach differs in motivation from similar attempts at identifying the propagation of errors from retracted papers [4], or identifying whether papers directly citing retracted papers should also be retracted (a retraction cascade) [26, 27]. Instead, we view weakening as a network-level problem, in which trustworthy and untrustworthy communities can be identified based on their use of retracted publications.

3.2.1. Next steps

The above outlined use of PageTrust has some clear limitations and leaves open questions about its proposed use case for tracking the propagation of distrust.

First, throughout this paper, we have assumed that all citations are of equal importance; however, it is clear that the importance of a citation differs by intent [28]. A citation may be influential (building upon cited work) or incidental (merely mentioning cited work, e.g., for context) [29]. Previous approaches to computing the propagation of errors from retracted publications have utilized citation intents [4]. In the future, we would like to extend our approach to *retraction as forgetting* with these enriched citation networks. However, we note that citation intent labels are challenging to collect and require the full-text of any scientific publications whereas our current approach relies only on network structure.

Similarly, we currently ignore retraction reasons when using PageTrust (see HASRETRACTIONREASON

in Table 1). However, we would expect that different retraction reasons confer different epistemic effects (e.g. data manipulation seriously affects the claims of the retracted paper). In future work, we plan to extend our implementation of PageTrust to account for retraction reasons.

An additional limitation of PageTrust is its computational cost. The computational cost of PageTrust grows with the mean degree of nodes in the network, the number of nodes with an outgoing negative edge (retracted papers), and the total number of nodes in the network. Therefore, it is not feasible to frequently update PageTrust scores for all papers in OpenAIRE.

4. Conclusion

To know when a work has been invalidated through retraction is critical for human researchers and research agents alike. Our conceptualization of retraction as forgetting enables two new views of retracted research data in OpenAIRE. For the retracted publication, with *ignorance* we get a distinct retraction warning, in the metadata or realized in the browser interface. Secondly, for every non-retracted paper, we provide a measure of *weakening* as a likeliness of being compromised by outgoing citation links that point to retracted publications. Both may be interpreted by a literature agent as the particular paper having lost its ability to support further research.

While it is likely just a matter of time until OpenAIRE includes information on retraction status, it remains open what its representation and corresponding mechanism will look like. A recent article by one of OpenAIRE's Communication & Social Media Specialists calls for *researchers who have already built upon, funded, or cited publications, in the event of a publication being retracted, to immediately reassess their own work and mitigate long-term risks*;²⁰ our approach to handling retractions enables this.

Acknowledgments

The authors would like to thank Jodi Schneider for advice and helpful comments.

Declaration on Generative AI

During the preparation of this work, the authors used Claude Fable 5 in order to: Peer review simulation and content enhancement (deep research, strictly for information retrieval). Further, the authors used Sonnet 5 for Figure 1: content enhancement (write data visualization code). After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] Wiley, ExplanAItions 2025: The evolution of AI in research., 2025. URL: <https://www.wiley.com/en-us/about-us/ai/study/key-findings>.
- [2] W. Zhang, X. Li, Y. Zhang, P. Jia, Y. Wang, H. Guo, Y. Liu, X. Zhao, Deep research: A survey of autonomous research agents, 2025. doi:10.48550/arXiv.2508.12752, preprint.
- [3] S. Labenbacher, M. Niederer, S. Hammer, M. Bader, N. Schreiber, H. Bornemann-Cimenti, Performance of AI tools in citing retracted literature: Content analysis, Journal of Medical Internet Research 28 (2026) e88766. doi:10.2196/88766.
- [4] H. Zheng, Y. Fu, M. J. Sarol, I. Sarraf, J. Schneider, Addressing unreliability propagation in scientific digital libraries, in: Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries, JCDL '24, Association for Computing Machinery, New York, NY, USA, 2025. doi:10.1145/3677389.3702526.

²⁰<https://www.openaire.eu/community/blogs/securing-the-knowledge-commons-how-open-science-fights-academic-fraud>

- [5] J. Brainard, Silence greets requests to flag retracted studies, *Science* 377 (2022) 11–12. doi:10.1126/science.add6988.
- [6] J. Schneider, D. Ye, A. M. Hill, A. S. Whitehorn, Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data, *Scientometrics* 125 (2020) 2877–2913. doi:10.1007/s11192-020-03631-1.
- [7] C. Candal-Pedreira, A. Ruano-Ravina, Retracted studies in systematic reviews and clinical guidelines, *BMJ* 389 (2025) r724. doi:10.1136/bmj.r724.
- [8] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, J. Schirrwagen, H. Dimitropoulos, S. La Bruzzo, I. Foufoulas, A. Mannocci, M. Horst, A. Czerniak, K. Iatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, A. Lempesis, A. Ioannidis, N. Manola, P. Principe, T. Vergoulis, S. Chatzopoulos, D. Pierrakos, OpenAIRE graph dataset, 2022. doi:10.5281/zenodo.7488618.
- [9] C. Atzori, A. Bardi, P. Manghi, A. Mannocci, The OpenAIRE workflows for data management, in: C. Grana, L. Baraldi (Eds.), *Digital libraries and archives*, Springer International Publishing, Cham, 2017, pp. 95–107. doi:10.1007/978-3-319-68130-6_8.
- [10] M. Rittman, Retraction Watch retractions now in the Crossref API, *Crossref Blog*, 2025. doi:10.13003/692016.
- [11] C. Hauschke, S. Nazarovets, (non-)retracted academic papers in openalex, *Journal of Information Science* 52 (2026) 993–997. doi:10.1177/01655515251322478.
- [12] C. D. Kerchove, P. V. Dooren, The PageTrust algorithm: How to rank web pages when negative links are allowed?, in: *Proceedings of the 2008 SIAM International Conference on Data Mining*, Society for Industrial and Applied Mathematics, 2008, pp. 346–352. doi:10.1137/1.9781611972788.31.
- [13] J. Priem, H. Piwowar, R. Orr, OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts, in: *Proceedings of the 26th International Conference on Science, Technology and Innovation Indicators (STI 2022)*, 2022. doi:10.5281/zenodo.6936227.
- [14] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, J. Schirrwagen, H. Dimitropoulos, S. La Bruzzo, I. Foufoulas, A. Mannocci, M. Horst, A. Czerniak, K. Iatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, A. Lempesis, A. Ioannidis, N. Manola, P. Principe, T. Vergoulis, S. Chatzopoulos, D. Pierrakos, OpenAIRE research graph dataset, 2019. doi:10.5281/zenodo.3516917.
- [15] T. Kuhn, R. Taelman, V. Emonet, H. Antonatos, S. Soiland-Reyes, M. Dumontier, Semantic micro-contributions with decentralized nanopublication services, *PeerJ Computer Science* 7 (2021) e387. doi:10.7717/peerj-cs.387.
- [16] J. Schneider, J. Lee, H. Zheng, M. O. Salami, Assessing the agreement in retraction indexing across 4 multidisciplinary sources: Crossref, Retraction Watch, Scopus, and Web of Science, *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)* (2023). doi:10.55835/6441e5cae04dbe5586d06a5f.
- [17] NISO Communication of Retractions, Removals, and Expressions of Concern Working Group, NISO RP-45-2024, Communication of Retractions, Removals, and Expressions of Concern (CREC), 2024. doi:10.3789/niso-rp-45-2024.
- [18] S. Peroni, D. Shotton, The SPAR ontologies, in: D. Vrandečić, K. Bontcheva, M. C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L.-A. Kaffee, E. Simperl (Eds.), *The Semantic Web – ISWC 2018*, Springer International Publishing, Cham, 2018, pp. 119–136. doi:10.1007/978-3-030-00668-6_8.
- [19] S. Peroni, D. Shotton, FaBiO and CiTO: Ontologies for describing bibliographic resources and citations, *Journal of Web Semantics* 17 (2012) 33–43. doi:10.1016/j.websem.2012.08.001.
- [20] S. Peroni, D. Shotton, F. Vitali, Scholarly publishing and linked data: Describing roles, statuses, temporal and contextual extents, in: *Proceedings of the 8th International Conference on Semantic Systems, I-SEMANTICS '12*, Association for Computing Machinery, New York, NY, USA, 2012, p. 9–16. doi:10.1145/2362499.2362502.
- [21] DataCite Metadata Working Group, DataCite metadata schema documentation for the publication and citation of research data and other research outputs, Documentation Version 4.7, DataCite e.V., 2026. doi:10.14454/qdd3-ps68.
- [22] C. Beierle, D. Brand, D. Howey, G. Kern-Isberner, K. Sauerwald, S. Todorovikj, M. Ragni, Intentional forgetting operators—Formal foundations and empirical support, Springer Nature Switzerland,

- 2026, pp. 129–157. doi:10.1007/978-3-032-17621-9_7.
- [23] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank citation ranking: Bringing order to the web, Technical Report, Stanford University, 1998.
 - [24] P. Gautret, J.-C. Lagier, P. Parola, V. T. Hoang, L. Meddeb, M. Mailhe, B. Doudier, J. Courjon, V. Giordanengo, V. E. Vieira, H. Tissot Dupont, S. Honoré, P. Colson, E. Chabrière, B. La Scola, J.-M. Rolain, P. Brouqui, D. Raoult, RETRACTED: Hydroxychloroquine and azithromycin as a treatment of COVID-19: Results of an open-label non-randomized clinical trial, *International Journal of Antimicrobial Agents* 56 (2020) 105949. doi:10.1016/j.ijantimicag.2020.105949.
 - [25] A. Bardi, I. Kuchma, G. Pavone, M. Artini, C. Atzori, A. Bäcker, M. Baglioni, A. Czerniak, M. De Bonis, H. Dimitropoulos, I. Foufoulas, M. Horst, K. Iatropoulou, A. Kokogiannaki, S. La Bruzzo, E. Lazzeri, P. Manghi, A. Mannocci, N. Manola, et al., [Dataset] OpenAIRE Covid-19 publications, datasets, software and projects metadata, 2023. doi:10.5281/zenodo.8221703.
 - [26] M. Usman, W.-T. Balke, On retraction cascade? Citation intention analysis as a quality control mechanism in digital libraries, in: O. Alonso, H. Cousijn, G. Silvello, M. Marrero, C. Teixeira Lopes, S. Marchesin (Eds.), *Linking Theory and Practice of Digital Libraries*, Springer Nature Switzerland, Cham, 2023, pp. 117–131. doi:10.1007/978-3-031-43849-3_11.
 - [27] M. Usman, W.-T. Balke, Tracing the retraction cascade: Identifying non-retracted but potentially retractable articles, in: A. Antonacopoulos, A. Hinze, B. Piwowarski, M. Coustaty, G. M. Di Nunzio, F. Gelati, N. Vanderschantz (Eds.), *Linking Theory and Practice of Digital Libraries*, Springer Nature Switzerland, Cham, 2024, pp. 109–126. doi:10.1007/978-3-031-72437-4_7.
 - [28] N. R. Aljohani, A. Fayoumi, S.-U. Hassan, An in-text citation classification predictive model for a scholarly search system, *Scientometrics* 126 (2021) 5509–5529. doi:10.1007/s11192-021-03986-z.
 - [29] M. Valenzuela, V. A. Ha, O. Etzioni, Identifying meaningful citations, in: *AAAI Workshop: Scholarly Big Data*, 2015. URL: <https://aaai.org/papers/aaaiw-ws0121-15-10185/>.
 - [30] P. Gautret, J.-C. Lagier, P. Parola, V. T. Hoang, L. Meddeb, M. Mailhe, B. Doudier, J. Courjon, V. Giordanengo, V. E. Vieira, H. T. Dupont, S. Honoré, P. Colson, E. Chabrière, B. La Scola, J.-M. Rolain, P. Brouqui, D. Raoult, Retraction notice to “Hydroxychloroquine and azithromycin as a treatment of COVID-19: Results of an open-label non-randomized clinical trial” [international journal of antimicrobial agents 56 (2020), 105949], *International Journal of Antimicrobial Agents* 65 (2025) 107416. doi:10.1016/j.ijantimicag.2024.107416.

A. Sample JSON-LD of a retracted publication and its retraction notice

```

1 {
2   "@context": {
3     "cito": "http://purl.org/spar/cito/",
4     "fabio": "http://purl.org/spar/fabio/",
5     "pso": "http://purl.org/spar/pso/",
6
7     // Properties of retracted research product
8     "isRetractedBy": {"@type": "cito:isRetractedBy"},
9     "hasRetractionDate": {"@type": "fabio:hasRetractionDate"},
10    "hasRetractionStatus": {"@type": "pso:holdsStatusInTime"},
11  },
12
13  "id": "doi_dedup___:0b1c...",
14  "@type": "fabio:Work",
15  "mainTitle": "Hydroxychloroquine and azithromycin as a treatment of COVID-19:
16               results of an open-label non-randomized clinical trial",
17  "instances": [
18    {
19      "pids": [{"scheme": "doi",
20               "value": "10.1101/2020.03.16.20037135"}],
21      "@type": "fabio:Preprint",
22      "publicationDate": "2020-03-20"
23    }
24  ]
25  // preprint: no retraction status

```

```

24 },
25 {
26   "pids": [{"scheme": "doi",
27             "value": "10.1016/j.ijantimicag.2020.105949"}],
28   "@type": "fabio:JournalArticle",
29   "publicationDate": "2020-07-01",
30   "retractionDate": "2024-12-16",
31   "isRetractedBy": "https://doi.org/10.1016/j.ijantimicag.2024.107416",
32   "retractionStatus": {
33     "@type": "pso:StatusInTime",
34     "pso:withStatus": "pso:retracted-from-publication",
35     "hasRetractionType": "Retraction",
36     "hasRetractionReason": [
37       "Ethical Violations by Author",
38       "Objections by Author(s)"
39     ]
40   },
41 },
42 // ...
43 ]
44 }
45
46 // The retraction notice, already in the graph as its own record,
47 // retrieved by a separate request; @context as above
48 {
49   "@context": {
50     "cito": "http://purl.org/spar/cito/",
51     "fabio": "http://purl.org/spar/fabio/",
52
53     // Properties of retracting research product (retraction notice)
54     "retracts": {"@type": "cito:retracts"},
55   },
56
57   "@type": "fabio:RetractionNotice",
58   "id": "https://doi.org/10.1016/j.ijantimicag.2024.107416",
59   "mainTitle": "Retraction notice to Hydroxychloroquine...",
60   // ...
61
62   "retracts": "https://doi.org/10.1016/j.ijantimicag.2020.105949",
63   "hasSource": "Publisher"
64 }

```

Listing 1: The proposed OpenAIRE record for Gautret et al. [24] and its retraction notice [30] with our schema extension marked in bold; some keywords omitted for readability.