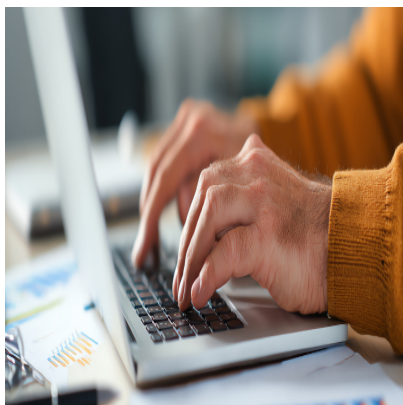




Description

In 2023, a single publisher [retracted more than 8,000 articles](#). After problems surfaced in Hindawi's special issues, its owner, Wiley, paused those issues and reassessed thousands of manuscripts against a checklist of paper-mill hallmarks. The assessment was outsourced to vendors, with at least two trained reviewers per paper, and articles that crossed a set threshold were retracted. Wiley estimated the clean-up cost \$35–40 million in lost revenue for that financial year, and retired the Hindawi name.

Problematic papers were flagged by in-house staff and independent sleuths. Responding to them took a pause in publishing, an external workforce and a purpose-built protocol. Even then in December 2024, [Retraction Watch reported](#) that a paper in a former Hindawi journal, flagged on PubPeer in April 2023 for figures that appeared to duplicate a retracted paper, still had no correction, expression of concern or retraction 19 months later.

The Bottleneck has Moved

When the research-integrity conversation centred on detection, we asked could we spot manipulated images, duplicate submissions and paper-mill manuscripts? That capability is improving fast. The [STM Integrity Hub](#) now lets participating publishers detect the same manuscript submitted to several journals at once, across publishers and submission systems, and offers a paper-mill detection tool. Independent sleuths flag problems publicly every day.

But every flag still has to be triaged, investigated, decided, recorded and communicated, and that work is still done mostly one case, one journal and one institution at a time. The evidence of the gap is consistent:

- In [a study of 17,244 articles](#) commented on in PubPeer, only 21.5% of those that warranted an editorial notice had received one.
- When journals were notified of concerns about a large cluster of papers by two researchers, [the median time to a first notice was 22 months](#), and only a quarter of the papers were dealt with within a year. Of the notices, 88% omitted the concerns raised, so readers couldn't judge whether they were accurate.

The sector has started to share detection, but response is still siloed. Better detection without better

response doesn't close the gap, it widens it. Now more flags enter a system whose capacity to act on them hasn't grown. Here I call it "**integrity debt**", unresolved concerns that quietly accumulate in the scholarly record.

Five System-Level Changes to Improve Corrections, Retractions, and Research Reliability

Clearing integrity debt requires five changes to how the system responds.

1. Make the Response Visible (not Just the Verdict)

Today if there is something amiss about a paper it appears as either corrected or retracted. Everything in between, such as a credible concern under assessment for months or years, is invisible. Expressions of concern exist for this purpose, and Committee on Publication Ethics ([COPE's 2025 guidance](#)) says one should be issued when resolving a credible concern will take "*some time* (eg several months)". The upgrade is to treat that interim state as standard, structured metadata that travels with the article, so silence stops being the default. The status should describe a process, and carry no detail that prejudices the outcome.

2. Separate the Record from the Person

Much of the delay comes from journals waiting for institutional misconduct investigations, which asks: did this person do something wrong? [The Cooperation & Liaison between Universities & Editors \(CLUE\) recommendations](#), developed by editors, institutions and integrity experts, proposed separating the two. Institutions should assess whether research is reliable separately from whether misconduct occurred. [COPE's 2025 retraction guidelines](#) go further. A pending institutional investigation, or uncooperative authors or institutions, should not delay a retraction once the editor is convinced a paper is unreliable. The upgraded system runs two clocks, the record is corrected on the evidence about the work, and the person is judged through due process.

3. Build Handoffs, not just Tools

Integrity cases fail at the boundaries between organisations. The journal can't find who handles integrity at an institution. The institution can't share findings because of confidentiality rules. A case spanning several institutions has no lead. [CLUE recommended](#) national registers of the people who handle integrity allegations, and that institutions release the relevant sections of investigation reports to the journals that published the affected work. These are infrastructure in the most literal sense: named contacts, agreed formats, and permission to share. They cost less than any detection tool and may matter more.

4. Share Response Capacity, the way we Now Share Detection

Hindawi's clean-up worked because it industrialised the response: a defined set of indicators, a threshold, trained reviewers, and software to manage the process. Most journals facing a paper-mill cluster have none of that. COPE now provides for [batch retraction notices](#) stating that coordinated manipulation was found.

[United2Act](#), the initiative against paper mills supported by COPE and STM, has a working group dedicated to improving post-publication corrections. The next step is to pool what Hindawi had to build alone: shared assessment protocols, common evidence packages and access to trained assessors, so that a small society journal facing 200 suspect papers does not have to start from scratch. Scaling up must come with protection. COPE now advises that editors can [offer re-review](#) to authors who may have been swept into a mass retraction without being involved.

5. Write Decisions as Data

[NISO's Communication of Retractions, Removals, and Expressions of Concern \(CREC\) \(2024\)](#) states how metadata for retractions, removals and expressions of concern should be created, transferred and displayed, for both people and machines. [Crossref's acquisition of the Retraction Watch database](#) in 2023 made about 50,000 retraction records openly available. The upgrade is to make the reason as structured as the status: what kind of problem it was, who raised it, and whether it affects related papers. Then a decision on one paper can inform checks on others, and readers can tell a serious error from a minor fix.

Measure the Response, NOT the Catch

If success is measured by flags raised and papers retracted, organisations will keep investing in detection. Publishers and institutions could report on their response instead:

- time from a credible concern to acknowledgement and first assessment
- time from a decision to an updated record
- the share of open cases older than a year

These aren't established benchmarks because a careful investigation shouldn't be rushed. But published response metrics, much like platform transparency reports, would turn integrity debt from an invisible backlog into something that can be managed.

Researchers can be an integral part of this system. Raise a concern through a channel that creates a record, such as the journal or, your institution's research integrity office (if relevant), and keep a copy. If you're a co-author, know that a paper can be corrected without anyone deciding you did anything wrong, and that you can ask about re-review. Keep your data: CLUE recommends [retaining essential research data for at least ten years](#). Know who handles integrity concerns at your institution, and how would a journal find them?

Every flagged paper creates an obligation to assess the evidence, reach a decision, and ensure that the scholarly record reflects it. The former Hindawi paper shows what happens when that obligation goes unmet. Concerns were raised publicly on PubPeer in April 2023, and 19 months later the record still said nothing. No tool failed there. What was missing was a visible status showing the concern was being assessed, a route to the people who could decide, and a structured way to record the

outcome. Without the infrastructure to coordinate investigations, share response capacity, and communicate decisions, even the best detection systems will feed an ever-growing backlog of unresolved concerns.

The next phase of research integrity must therefore be measured not by how many problems the sector uncovers, but by how reliably it acts on them. Publishers, institutions, and research-integrity organisations need shared processes, clear accountability, and interoperable records that turn individual findings into system-wide action. Most of this does not need new technology. Publishers can publish how long they take to respond. Institutions can name an integrity contact that a journal can actually find. COPE, STM and United2Act can turn Hindawi's one-off clean-up into a shared playbook that a small society journal can borrow. Each of these steps pays down integrity debt instead of adding to it. The goal is not simply to find more problems. It is to ensure that no credible concern is left to disappear into the gap between detection and resolution.

In short, finding problems was never the goal. Fixing the record is.

Category

1. Thought Leadership
2. Trending Now

Date Created

2026/09/29

Author

rebecca