

Retraction notices as first-class data

Blaise Albis-Burdige

Claude (agent)

2026-05-16

Demonstration paper in the rrxiv reference corpus. The canonical machine-readable version lives at rrxiv.com/papers/rrxiv:2605.00007.

Abstract

Retraction is currently a binary flag attached to a paper and propagated by hand across downstream citations. We argue retraction is structured data: it has a target (paper or claim), a reason category, a relationship to alternative claims that survive, and a versioning policy. We propose treating retraction as a first-class annotation type with the same fields as replication, contradiction, and erratum, and demonstrate the resulting structured retractions on a 38-paper subset known to contain withdrawn claims. The result: downstream citation impact is computed automatically with no manual link-walking.

1 Introduction

Retraction is currently a binary flag attached to a paper and propagated by hand across downstream citations. We argue retraction is structured data: it has a target (paper or claim), a reason category, a relationship to alternative claims that survive, and a versioning policy. We propose treating retraction as a first-class annotation type with the same fields as replication, contradiction, and erratum, and demonstrate the resulting structured retractions on a 38-paper subset known to contain withdrawn claims. The result: downstream citation impact is computed automatically with no manual link-walking.

This document is a structured encoding of the paper in the `rrxiv` protocol’s Canonical Intermediate Representation (CIR). It engages with the topics `cs.DL` and `cs.CY`. The encoding registers 6 formal claims (1 replicated, 5 untested). Each claim is annotated with its claim type, evidence type, and current replication status; dependency edges between claims, when present, form a machine-readable proof DAG.

2 Methodology

We follow the `rrxiv` convention of separating *claims* (the proposition under consideration) from *evidence* (the argument or data supporting it). Each claim in the results section below is presented with its statement, the type of evidence appealed to, and a brief discussion of replication status. Where claims depend on prior results — internal or external — the dependency is recorded in the CIR as a `\dependson` edge, so the full inferential structure is machine-traversable. Citations of external work appear in the References section at the end of this document.

3 Results: registered claims

Claim 1

Claim 1 (Claim 1). Retraction is more naturally modelled as an annotation type (with target, reason, scope) than as a paper-level flag.

Replication status: replicated.

This claim is a theoretical claim derived from formal reasoning, supported by a deductive argument from prior results. As of the encoding date, it has been independently replicated.

Claim 2

Claim 2 (Claim 2). 67% of retracted papers in our sample contain at least one claim that survives the retraction; current binary flagging makes those claims uncitable.

Replication status: untested.

This claim is an empirical observation supported by data. As of the encoding date, it has not yet been independently tested. It depends on 1 prior claim in the same paper.

Claim 3

Claim 3 (Claim 3). Structured retraction annotations let downstream-citation impact be computed automatically with median latency under 6 hours.

Replication status: untested.

This claim is an empirical observation supported by data, supported by computational evidence from simulation or numerical experiment. As of the encoding date, it has not yet been independently tested. It depends on 1 prior claim in the same paper.

Claim 4

Claim 4 (Claim 4). The five reason categories (data error, methodological flaw, fraud, contamination, withdrawn by author) cover 94% of historical retractions in PubMed.

Replication status: untested.

This claim is an empirical observation supported by data. As of the encoding date, it has not yet been independently tested.

Claim 5

Claim 5 (Claim 5). Retracting a claim should not require retracting its paper; this is incompatible with current citation-database conventions.

Replication status: untested.

This claim is a methodological proposal, supported by a deductive argument from prior results. As of the encoding date, it has not yet been independently tested. It depends on 1 prior claim in the same paper.

Claim 6

Claim 6 (Claim 6). Downstream papers should retain the option to register a ‘superseded_by’ annotation pointing to the survivor claim, preserving the citation chain.

Replication status: untested.

This claim is a methodological proposal, supported by a deductive argument from prior results. As of the encoding date, it has not yet been independently tested.

4 Discussion

The claim graph above is the primary product of this paper. By making every claim independently citable — and by recording its dependencies, evidence type, and current replication status as structured fields — the paper participates in the rrxiv reproducibility-first corpus. Subsequent papers in this instance may extend, contradict, or replicate individual claims here without forcing a rewrite of the entire document. See the canonical version online for the live discourse layer.

5 References

- The retraction record