

The Missing Piece: From Open Papers to Open Process

By: Sona Bhatia

The Gap

Open science was built to solve a trust problem: too many scientific claims, not enough transparency into how those claims were produced. The scale of that problem is well documented: more than 70% of researchers report failing to reproduce another scientist's work, and nearly 90% believe science faces a reproducibility crisis. Open-access publishing, data repositories, and preprint servers have improved access to scientific knowledge, but they still emphasize the final publication rather than the research process itself.

Artificial intelligence has introduced a new dimension to this challenge. Large language models can now summarize literature, draft manuscripts, assist with systematic reviews, generate code, and even help write peer reviews in minutes. Many publishers have responded by requiring authors to disclose AI use, but disclosure is not verification. A checkbox relies entirely on self-reporting, with no mechanism to determine whether AI was used responsibly, whether its outputs were independently verified, or whether the declaration is even truthful. It rewards honesty while offering little deterrence to misconduct.

The greater risk is not undisclosed AI use; rather, it is unverified AI use. A researcher may openly acknowledge using AI while unknowingly publishing hallucinated citations, unsupported claims, statistical errors, or inaccurate summaries. In healthcare and clinical research, these mistakes can propagate through systematic reviews, influence clinical guidelines, and ultimately affect patient care. Once incorrect information enters the scientific record, it can be cited repeatedly long before it is corrected.

Open science has successfully made publications more accessible. Its next challenge is making the scientific process itself transparent and verifiable, which is where this proposal begins.

Process Provenance

Rather than asking researchers to declare how they worked, open science should allow them to demonstrate how they worked. I propose Process Provenance: a standardized, evidence-based record of the research journey that accompanies every publication. Instead of evaluating only the finished manuscript, the scientific community gains visibility into how that manuscript evolved, what evidence supported it, and how AI-assisted work was verified by humans.

How TeraOpenScience Could Enable This

TeraOpenScience is already designed to connect researchers, AI-powered tools, discussion spaces, manuscript workflows, and interdisciplinary collaboration. Process Provenance extends that ecosystem by treating the research process itself as a scientific artifact rather than hidden background work. Instead of becoming another disclosure form, TeraOpenScience could become the platform where transparency is generated automatically throughout a project's lifecycle, built from six connected pieces:

Living research timelines. Each project maintains an automatically generated timeline, manuscript version history, methodology revisions, dataset versions, analysis updates, AI-assisted interactions, contributor activity, linked discussions, and decisions. Rather than reconstructing a workflow at submission, the platform captures it continuously as research progresses.

AI-assisted provenance reports. AI shifts from writing assistant to transparency assistant. Using the project's history, TeraOpenScience could automatically generate a reproducibility report summarizing datasets used, preprocessing steps, software versions, model parameters, statistical analyses, references requiring verification, and AI-generated content still needing human confirmation, reducing administrative burden while making reproducibility easier.

Evidence rather than disclosure. Human research evolves gradually through revisions, corrections, and experimentation; AI-generated work simply copied into a manuscript leaves a very different pattern. Reviewers wouldn't attempt to "detect AI." They'd examine whether the research record reflects a genuine process supported by evidence, shifting the question from 'was AI used' to 'Was every conclusion independently validated? '

Intelligent random audits. Not every submission requires exhaustive review. Much like financial auditing, TeraOpenScience could randomly select a percentage of projects for deeper verification, examining the provenance report and supporting evidence rather than searching for AI itself. Inconsistencies could trigger public correction notices, reproducibility flags, institutional notification, or retraction. The objective isn't perfect detection but credible accountability.

Rewarding good science. Transparency shouldn't exist only to catch misconduct. Researchers who consistently document reproducible workflows could earn reproducibility badges, verified workflow certifications, open methodology awards, and trusted researcher status, giving students and early-career researchers tangible, visible evidence of rigorous practice.

Privacy by design. Clinical datasets, proprietary collaborations, and sensitive patient information stay protected through access controls. Rather than exposing confidential material, the platform stores cryptographic hashes, timestamps, metadata, and audit logs that verify a process occurred while preserving privacy and intellectual property, demonstrating integrity without revealing what can't ethically or legally be shared.

Why Healthcare AI Needs This Most

Healthcare AI depends on trust. Researchers increasingly use AI to summarize evidence, analyze datasets, generate code, and assist with literature reviews, and these tools accelerate discovery, but they also raise the chance that inaccurate information enters medical research unnoticed. A hallucinated citation today can become tomorrow's systematic review, next year's clinical guideline, and eventually a factor in patient care. Making AI-assisted healthcare research transparent is therefore not merely academic integrity; it is a patient safety issue. Process Provenance ensures every AI-assisted conclusion carries documented human verification before it enters the scientific record.

Expected Impact

Implementing Process Provenance through TeraOpenScience would ripple across the research ecosystem: researchers spend less time recreating documentation because provenance is generated continuously; reviewers receive structured evidence instead of relying on disclosure statements alone; institutions gain stronger confidence in published work's integrity; students receive visible credit for iterative work that usually goes unseen; and clinicians and policymakers can trust AI-assisted research because the reasoning behind it is transparent and independently verifiable. Most importantly, transparency becomes proactive rather than reactive.

Conclusion

Open science has made publications more accessible, but accessibility alone cannot guarantee trust. The next evolution is not simply sharing more papers; it is sharing the evidence behind how those papers came to exist. The missing piece is Process Provenance: transforming the scientific process itself into a transparent, verifiable, collaborative record. By embedding AI-assisted provenance tracking, reproducibility reporting, and evidence-based verification directly into TeraOpenScience, the platform could move beyond asking researchers to promise integrity, and instead help them demonstrate it. Open science should not only make knowledge visible, but it should also make the path to knowledge visible too.