

Sources Should Travel With Claims

A citation can identify a source without preserving the evidentiary relationship.

DRAFT PREVIEW DENNIS HEDEGREEN ANALYSE DRAFT-LOCAL

[HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/SOURCES-SHOULD-TRAVEL-WITH-CLAIMS/](https://hedegreenresearch.com/articles/sources-should-travel-with-claims/)

A citation can identify a source without preserving the evidentiary relationship.

A sentence ends with [1].

The reader can click it.

Maybe it opens a paper. Maybe it jumps to a source list. Maybe it points to a government report, a company filing or another article.

That is useful.

But it still leaves a basic question unanswered:

Why is this source here?

A citation can identify a source without preserving the evidentiary relationship between that source and the claim beside it.

Journalism is not only a collection of sources.

It is a set of claims connected to evidence.

Those connections should survive publication.

A source list tells you what was read

Consider a sentence:

> Country X cut emissions by 40% while GDP grew by 20%.

One source may support the emissions figure.

Another may support the GDP figure.

A third may define the period.

A fourth may explain whether emissions are territorial, consumption-based or net of land-use changes.

A bibliography containing all four sources is useful.

But the bibliography alone does not tell the reader which source supports which part of the sentence.

That is the difference between a source list and provenance.

> **A source list tells you what was read. Provenance tells you what supports what.**

The important relation is not only:

article → source

It is:

claim → source

And sometimes:

claim → source → role → reason

The role matters because sources do different jobs.

One source may originate a claim.

Another may independently verify it.

Another may provide data.

Another may contradict it.

Another may define a method without supporting the conclusion.

Flattening all of those into one reference list throws away structure the newsroom already understands.

Why this source is here

That structure does not need to dominate the page.

A citation such as [1] can stay compact.

But when a reader chooses to inspect it, the citation could reveal something like:

Vernor Vinge — 1993 *The Coming Technological Singularity*

Why this source is here Supports the interpretation that Vinge framed the singularity as a boundary beyond which existing models for predicting the future become unreliable.

Role Primary source / supports interpretation

The important field is not the title.

It is:

Why this source is here.

That explanation should not summarize the entire source.

It should answer a narrower question:

> Why does this source belong to this claim?

This becomes especially useful when a source supports only one part of a compound sentence, provides data but not interpretation, or originates a claim without independently verifying it.

The relation itself is editorial metadata.

It can be wrong.

It should be reviewable.

It is part of the reporting record, not an oracle.

A URL is a location. It is not preservation.

There is another problem with treating citations as destinations.

The web changes.

A 2021 Harvard Law School study examined external hyperlinks in 553,693 New York Times articles published from 1996 through mid-2019.

Among deep links to specific pages, 25% were inaccessible.

Fifty-three percent of articles containing deep links had at least one rotted link.

But dead links were only part of the problem.

The researchers manually reviewed 4,500 links that were still technically reachable. Thirteen percent had drifted significantly from the content relevant when the Times article was published.

A link can therefore fail in two different ways.

> **Link rot breaks access. Content drift breaks meaning.**

A URL may still resolve while no longer showing the evidence the journalist relied on.

Pew Research Center found a similar contemporary fragility in 2024: 23% of sampled news webpages contained at least one broken external link.

The newsroom problem is not that webpages should never change.

It is whether the reporting record can reconstruct what the source was when it was used.

Preserve the source state used

Not every source needs the same preservation method.

A versioned journal article with a DOI may already have strong identity.

Legislation may have an official version or publication number.

A dataset may have a release identifier.

A content-addressed repository may already preserve version identity.

A live webpage is different.

So is a dashboard that changes daily.

A social post can disappear.

An online document can be edited without a visible history.

Those sources may require a snapshot, downloaded file, screenshot, timestamp, hash or archival capture.

The rule should not be:

> Snapshot everything.

It should be:

> **Preserve enough to identify the source state you relied on.**

The method depends on the source.

The requirement does not.

Preservation protects provenance, not truth

There is an important limit.

Perfectly archiving a bad source does not make it a good source.

A preserved press release remains the organization's own claim.

A downloaded spreadsheet can still contain bad data.

A captured webpage can still make a false assertion.

Preservation answers:

- what source was used? - in what version? - for which claim? - for what stated reason?

It does not, by itself, answer:

- is the source reliable? - is the method sound? - is the claim true?

Those remain verification and editorial questions.

> **Preservation protects provenance, not truth.**

Preservation is not publication

There is also a separate access problem.

A newsroom may need to preserve a source internally without necessarily having the right to republish the entire source publicly.

That distinction can matter for copyrighted reporting, paywalled material, licensed databases and other restricted sources.

The provenance record and the preserved copy do not need the same access rules.

A reader may be able to see the source identity, retrieval date, version information and reason it supports the claim without being able to download the newsroom's preserved copy.

> **Preservation and publication are different acts.**

The exact legal boundary depends on jurisdiction and source type and requires separate review.

This is a design distinction, not a claim that every source may be freely archived and republished.

Rendering should not destroy provenance

The problem does not end at the webpage.

Articles are exported.

PDFs are downloaded.

Research is reused.

Claims are summarized by machines.

A finished document should not become the dead end of the evidence that created it.

Hedegreen Research is experimenting with one implementation:

`article.pdf`

plus

`article.json`

The PDF is the human-readable rendering.

The JSON can preserve the structured relationships underneath it:

- claim identifier; - source identifier; - source role; - reason for the relation; - source-version information; - retrieval time.

That is one implementation, not the standard.

Another publisher might use structured HTML, embedded metadata, archival packages or an API.

The deeper rule is:

> **Rendering should not destroy provenance.**

A web citation, a PDF footnote and a machine-readable record can be different representations of the same underlying evidence relationship.

Store more provenance than you show

Readers do not need to see every field.

Most people will never inspect every citation.

That is fine.

The visible article should remain readable.

The newsroom can preserve more structure underneath.

A citation marker may be enough for most readers.

A hover or tap can expose the evidentiary rationale.

A deeper source record can exist for someone who wants to audit the claim.

> **Store more provenance than you show.**

That avoids turning every article into a database interface while still keeping the evidence inspectable.

Three principles

A minimal standard could be:

1. **Bind evidence to the claim.** Preserve which source is relevant to which claim, what role it plays and why.
2. **Preserve the source state used.** Keep enough version information to reconstruct what the newsroom actually relied on.
3. **Carry provenance forward.** Do not lose the relationship when the story is rendered, exported, updated or archived.

No universal citation widget is required.

No universal JSON format is required.

No universal archive is required.

The important thing is that a claim should retain the identity, role and version of the evidence used to support it.

Implementation status at Hedegreen Research

Hedegreen Research is actively experimenting with structured citation objects, source preservation, and PDF + JSON export.

The specific reader-facing citation popover described here — including a **“Why this source is here”** explanation — should be treated as a proposed/experimental interface unless it is already deployed in the current production article system at publication time.

The article's standard does not depend on that exact interface.

What is still unproven

Several questions remain.

How much provenance should readers see by default?

Which relation labels help rather than create bureaucracy?

How should a newsroom handle a source that is corrected or retracted?

How should compound claims be divided?

What legal rules apply to preservation in different jurisdictions?

How should provenance travel once content leaves the publisher's control?

Those questions need testing.

But the underlying problem is already measurable.

A citation can survive while the evidentiary relationship disappears.

A URL can survive while the cited content changes.

A PDF can survive while the source structure that created it is lost.

That should not be the default.

A source list tells you what was read.

Provenance tells you what supports what.

And a finished document should not be the end of that chain.

RELATION MEMORY

NEARBY [Claims Need States](#) · Previous instalment in Journalism, Reassembled. Article 004 preserves claim wording, source position and verification state; Article 005 preserves the evidence relationship attached to the claim.

NEARBY [Questions Should Not Disappear](#) · Both articles preserve context that ordinary article prose often loses: open questions in Article 003, claim-source relationships in Article 005.

NEARBY [A Percentage Needs a Reference Frame](#) · Both articles argue that a visible statement needs its evidentiary frame to survive publication and reuse.

CANONICAL URL: [HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/SOURCES-SHOULD-TRAVEL-WITH-CLAIMS/](https://hedegreenresearch.com/articles/sources-should-travel-with-claims/)
PDF VIEW GENERATED: 2026-09-29T10:12:08.044Z