

The Flow Asymmetry Audit — Methodology

Companion to the Dispersal Hypothesis · how AI-measured "flow" turns a 130-year argument into a test · Cryptographic Audit series · August 2026

The claim under test. By the Dispersal Hypothesis, Owen's *Sir Francis Bacon's Cipher Story* is the ORIGINAL narrative, dispersed by its author in fragments across the First Folio and other era works. If true, the direction of composition left a measurable fingerprint: the reassembled original reads **smoothly**; the host works read **peculiarly** at the insertion seams. If false — if Owen authored the story by splicing — his books inherit the seams instead. Either way, the numbers speak.

Color code used throughout: ■ Owen's text (claimed original) ■ host works (Folio & others) ■ guide words ■ claimed seams ■ controls / null corpora

1 · The dispersal picture

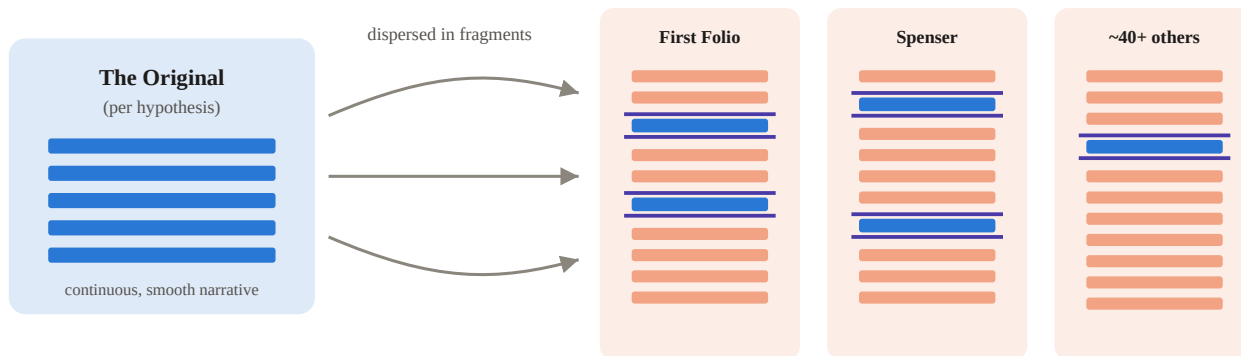


Figure 1 — The dispersal picture. Blue lines are the claimed original narrative; orange columns are host works; each inserted blue fragment sits between violet seams. The hypothesis predicts the blue text flows smoothly when reassembled, while the orange text stutters precisely at the violet boundaries.

2 · The two measurements

Measurement A — the source side. Score the internal flow of Owen's five volumes (the OCR'd corpus) with a language model: windowed perplexity and sentence-to-sentence coherence over the whole text. Compare against (i) the same text with paragraphs shuffled, and (ii) length-matched Victorian prose controls. *Prediction if the hypothesis holds:* Owen's text scores as coherent, connected narrative — despite being assembled from fragments of dozens of works, which is the surprising part a skeptic must explain.

Measurement B — the host side. In the First Folio (and later, each source work the canvas catalog identifies), score local flow in sliding windows centered on claimed extraction sites, versus windows centered on matched sites elsewhere in the same play (same scene-type, same speaker-density). *Prediction if the hypothesis holds:* a systematic flow anomaly concentrated at the claimed sites; *prediction if not:* no difference from the matched controls.

3 · What the instrument sees (ILLUSTRATIVE SKETCHES — not measured data)

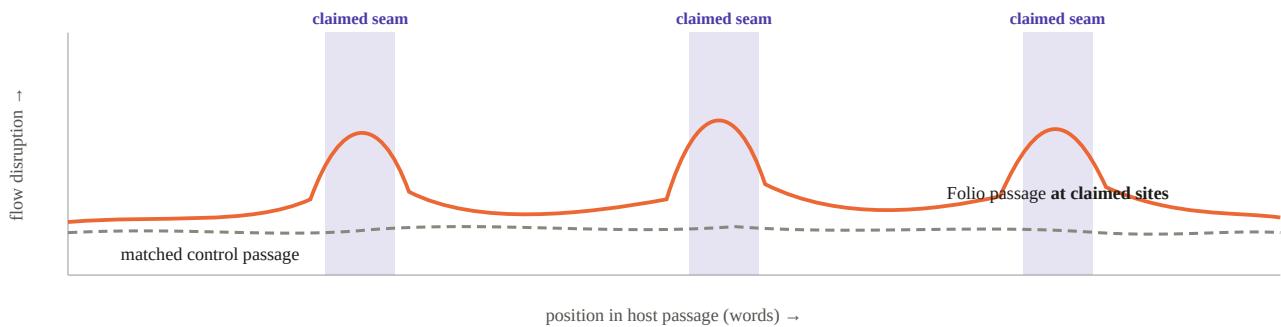


Figure 2 — The seam trace (illustrative). The orange curve is the flow-disruption score along a host passage; the predicted signature is peaks confined to the violet seam bands, while the gray dashed control passage stays flat. If the real data shows the orange curve flat too, the hypothesis fails this test — and that result is equally publishable.

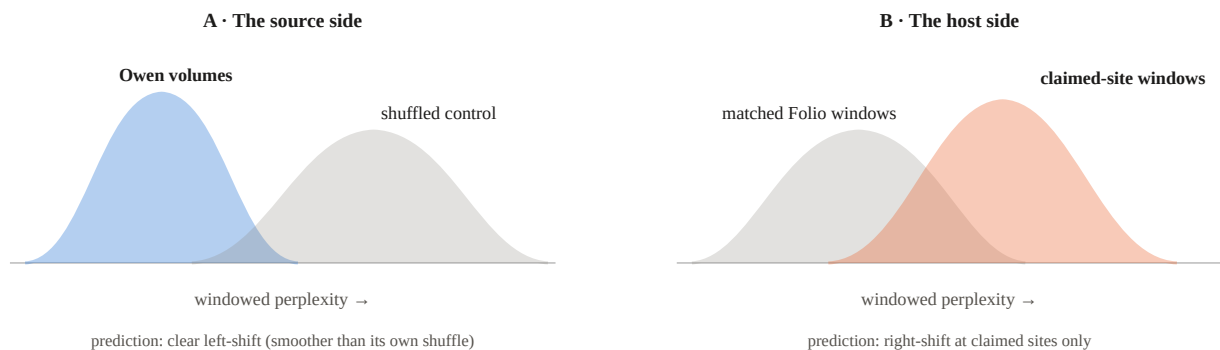


Figure 3 — The predicted asymmetry (illustrative). Two small multiples, one shared measure. Panel A: Owen's volumes (blue) should sit left of their own shuffled control (gray). Panel B: Folio windows at claimed extraction sites (orange) should sit right of matched windows (gray). The hypothesis needs BOTH shifts; either one alone is explainable by mundane causes.

4 • Metrics

Metric	What it measures	How
Windowed perplexity	How predictable the text is to a language model — the numerical form of "flow"	Fixed-size sliding windows (e.g., 50-word, 25-word stride); modern LM scored on period-normalized spelling; per-window score
Local coherence	Whether adjacent sentences belong together	Embedding similarity between consecutive sentences; drops mark discontinuity
Seam z-score	Whether disruption concentrates at claimed boundaries	Each claimed site's window score standardized against the distribution of matched-control windows from the same work
Guide-word clustering	Whether key words (Fortune, Honour, Nature...) cluster beyond chance	Observed spacing distribution vs. permutation null within the same work

5 • Null corpora & the guards

- **Nulls for Measurement A:** paragraph-shuffled Owen; sentence-shuffled Owen; length-matched Victorian narrative prose (period-fair baseline).
- **Nulls for Measurement B:** matched windows from the same play (same scene-type); randomized pseudo-sites drawn with the same guide-word density; other Jacobean plays outside the claimed corpus.
- **Circularity guard:** the Barsi-Greene masks list may *predict* which works carry fragments; it is never used to *define* the null corpora. The canvas catalog decides membership; the masks list is graded against it afterward.
- **OCR guard:** Books I–II are OCR output; all Measurement A scoring runs on human-spot-checked stretches, and any window overlapping an OCR artifact is dropped, not repaired.
- **Period guard:** Early Modern English scores as "peculiar" to a modern LM everywhere. All comparisons are therefore *within-work* and *within-period*: Folio-vs-Folio, Owen-vs-shuffled-Owen. No cross-era comparison is ever the evidence.

6 • Decision criteria — stated in advance

Outcome	Reading
Both predicted shifts appear (A left, B right), effect sizes reported with CIs	The Dispersal Hypothesis survives its hardest available test; proceed to the canvas-anchored phase
A appears, B does not	Owen's books are coherent but hosts show no seams — consistent with authored-not-recovered; the hypothesis is disconfirmed in its strong form
B appears, A does not	Anomalies at claimed sites but a choppy "original" — points to selection artifacts; investigate guide-word density confound before any stronger claim
Neither appears	Null result; published as such

Commitment: criteria precede data. Whatever the numbers say is the result — this is what makes it an audit rather than an argument.

7 • Answering the standing objections

The method pipeline's skeptical panel (after the Friedmans) makes three charges. This design answers each on its own terms:

Objection	The audit's answer
"No deterministic rule fixes which occurrence, how much text, or the order"	True of Owen's <i>description</i> — but the flow test doesn't require his rule. It tests the <i>product</i> : dispersal leaves seams whether or not the recovery rule was ever written down.
"The message is authored, not recovered"	That is exactly what Measurement A + B discriminate: an authored splice makes Owen's books carry the seams; a recovered original makes the hosts carry them. The objection becomes one branch of the decision table.

"Not reproducible — a second decoder gets a different story"

Deferred to the Reproducibility Core (Millett's 1895 self-demonstrating article is the known-plaintext test). Flow asymmetry is measurable regardless.

8 · Corpus status & run order

1. **Ready now:** Owen Books I–II + Synopsis (OCR'd, in Kaggle/); Books III–V (clean embedded text in Downloads — extract, minutes); First Folio (public domain).
2. **Phase F1:** Measurement A on the five volumes. No permissions needed.
3. **Phase F2:** Measurement B on the Folio using Owen's published citations as the claimed-site list.
4. **Phase F3 (after the scan):** re-run B with the canvas catalog's actual fragment list and colors — the definitive version, on the artifact's own testimony.

Prepared for Walter H. Wilkinson · gorhambury.org · The Flow Asymmetry Audit, Methodology v1 · August 2026 · Figures 2–3 are illustrative sketches of predicted signatures, not measured data · color palette validated for color-vision accessibility