A TECHNOLOGICAL AND INTELLECTUAL HISTORY

History of Diagnostics, Observability, and Debugging in Writing and Publishing

Textual evidence and the examination of change

Seven parts and 35 chapters

Including a chapter on Lyn Dupré's BUGS in Writing, Revised Edition

Concept by Dmitry Vostokov, DumpAnalysis.org

Developed with AI assistance (GPT-6 Astra Max and Fable 5.1 Extra) and checked against the sources listed in this article.

Version 6 | November 2026

Historical synthesis with 57 references,
chronology, glossary, and evidence taxonomy

Abstract

Writing and publishing leave evidence of their own construction. A deleted phrase, a corrected proof, a changed reference, and a failed document conversion preserve different aspects of that construction. In this history, we examine how writers, scribes, editors, printers, scholars, and software developers made such evidence available for analysis. We follow the investigation of textual discrepancies from manuscript correction and the printing house to word processing, digital publication, and AI-assisted composition.

Our main distinction is between the intended communication, the text that records it, the particular document that presents it, and the process that produces and distributes that document. A sentence can be grammatically acceptable while giving the wrong instruction. A correct manuscript can produce an unreadable page. A repaired publication can coexist with an earlier copy that continues to circulate. Each condition calls for a different comparison and a different kind of evidence.

Seven parts and 35 chapters connect historical cases with the practical examination of writing and publishing. A dedicated chapter considers Lyn Dupré's BUGS in Writing, Revised Edition: A Guide to Debugging Your Prose. Later chapters relate the history to our pattern-oriented software diagnostics and develop a worked investigation of a publication whose outputs disagree. A selected chronology, glossary, and evidence taxonomy provide additional viewpoints.

The history shows that correction becomes more dependable when the object, reference state, evidence, and consequences of a change remain distinguishable. Tools enlarge what we can inspect. They do not remove the need to decide what a discrepancy means, whose intention matters, or whether a proposed repair preserves the work.

How to read this history

Let us first specify the scope. Writing includes the construction and revision of literary, scholarly, technical, and other expository texts. Publishing includes editorial selection, preparation, production, distribution, correction, and preservation. We consider manuscripts, printed books, periodicals, electronic documents, and publications assembled by software. The examples are selective. They do not constitute a complete history of every language, writing system, or publishing tradition.

Diagnostics concerns the investigation of a condition and its possible explanations. Observability concerns the arrangements through which relevant states and changes can be inspected. We use this latter word comparatively for drafts, proofs, editorial correspondence, revision histories, and production records. We do not assert that an entire writing practice satisfies a mathematical definition of observability.

Debugging has several relationships to the subject. It is a modern comparison when applied to scribal correction or the repair of a printed sentence. It is an explicit prose metaphor in Dupré's book. It becomes literal when a conversion program, publishing script, font-processing component, or reading application contains a software defect. These relationships should not be collapsed into a claim that earlier practitioners were software debuggers or that their practices followed one continuous technological lineage.

Part I examines revision, textual witnesses, and the authority of a correction. Part II follows the printing house, errata, publication control, and material bibliography. Part III concerns composition, style, readability, and prose debugging. Part IV separates editorial responsibilities and kinds of verification. Parts V and VI examine computational production and the distributed life of a publication. Part VII develops the diagnostic synthesis.

Numbered references support historical and technical claims. Original publication dates are distinguished from later editions and digital representations. Standards and software documentation describe the versions or capabilities cited; they do not establish that an earlier publisher had those facilities. Examples introduced as hypothetical are analytical constructions. Connections to our analysis patterns are proposed applications, not claims of historical influence.

We retain the distinction between error and variation throughout. A nonstandard form may be deliberate. An authorized revision may change the meaning. A corrected spelling can destroy evidence needed by a textual scholar. Before examining how to repair a document, we therefore ask what the document is supposed to preserve.

Contents

Seven parts and 35 chapters

SECTION	PAGE
Abstract	2
How to read this history	2
PART I. MANUSCRIPTS REVISION AND TEXTUAL EVIDENCE	6
1. Vocabulary and historical scope	6
2. Ancient revision and the distance needed to reread	7
3. Manuscript correction and successive hands	7
4. Printing traditions beyond a European starting point	8
5. Textual criticism and the choice of a reference text	9
PART II. THE PRINTING HOUSE AND THE PUBLISHED RECORD	11
6. Composition and correction in the printing house	11
7. Proofs and the visibility of production work	11
8. Errata cancels and the persistence of an error	12
9. Publication control and the authority to alter	13
10. Material bibliography and the comparison of copies	14
PART III. COMPOSITION STYLE AND PROSE DEBUGGING	15
11. Drafts and the reconstruction of writing	15
12. Revision as a recursive activity	15
13. Style manuals and shared conventions	16
14. Readability measures and their limits	17
15. BUGS in Writing and the explicit debugging metaphor	18
PART IV. EDITORIAL WORK AND THE VERIFICATION OF CLAIMS	19
16. Structural editing copyediting and proofreading	19
17. Terminology navigation and document structure	20
18. Fact checking references and the evidence behind a sentence	20
19. Peer review and the limits of an editorial judgment	21
20. Translation and the conditions of a correct representation	22

SECTION	PAGE
PART V. COMPUTERS AND THE PRODUCTION OF DOCUMENTS	24
21. Mechanization and editable text	24
22. Spelling grammar and the Writer’s Workbench	25
23. Typesetting programs and the separation of source from output	25
24. Structured text and machine checkable relationships	26
25. Preflight color proofs and the physical publication	27
PART VI. VERSIONS DISTRIBUTION AND PRESERVATION	29
26. Revision histories and the identity of a version	29
27. OCR and the difference between an image and its text	29
28. Electronic books accessibility and multiple reading conditions	30
29. Metadata identifiers and distributed copies	31
30. Corrections retractions and preservation events	32
PART VII. DIAGNOSTIC SYNTHESIS AND WORKED INVESTIGATIONS	33
31. AI assisted writing and plausible incorrectness	33
32. Comparing evidence across the publication process	34
33. A repertoire of publishing analysis patterns	35
34. A worked investigation of conflicting publication outputs	36
35. Designing publications for future diagnosis	37
Conclusion	39
Appendix A Selected chronology	40
Appendix B Glossary	42
Appendix C Evidence taxonomy and blind spots	44
References	46

PART I

Manuscripts revision and textual evidence

1. Vocabulary and historical scope

Let us consider a book in which a reference points to the wrong chapter. Several explanations are possible. The author may have entered the wrong number. A chapter may have moved. A generated field may not have been updated. An export may have retained an earlier display value. The reader encounters one symptom, but the investigation concerns several possible stages of production.

We call a preserved item used in such an investigation a diagnostic artifact. It may be a manuscript leaf, a proof, an editorial query, an image of a page, a build log, or the actual file received by a reader. These artifacts have different relationships to the publication. A screenshot records an appearance. A source document records an editable representation. Neither necessarily records the settings that connected the two.

Correction requires a reference condition. For a quotation, that condition may be a particular source edition. For a chapter reference, it may be the structure of the approved manuscript. For literary dialogue, it may include a deliberate departure from ordinary grammar. A rule can help us detect a discrepancy without settling whether the discrepancy is a defect. The same unusual sentence may be an error in an instruction and a successful feature of a character’s speech.

The history also requires attention to authority. A scribe, author, publisher, censor, reader, and later editor can each alter a text for different reasons. The existence of an alteration does not tell us which person authorized it or which version a present edition ought to reproduce. We need evidence about the intervention as well as evidence of its visible result.

Our pattern-oriented viewpoint is concerned with recurring relationships among artifacts and explanations. We can compare versions, follow the history of a term, locate an interruption in a production sequence, or distinguish an observed failure from a proposed cause. These operations are useful across media, but their application depends on the material and institutional setting. A printing forme and a software build have comparable diagnostic questions without being the same kind of process.

The chapters that follow therefore examine both continuity and difference. Some problems recur because texts are copied and interpreted. Others arise from particular technologies. We can recognize a familiar pattern while still asking which mechanism could have produced it in this case.

2. Ancient revision and the distance needed to reread

Revision was an explicit subject of rhetorical teaching long before mechanical or electronic publishing. In Book 10, chapter 4 of the Institutio oratoria, Quintilian discusses addition, removal, and alteration. He recommends returning to writing after an interval, where circumstances permit, and also warns against excessive correction. The passage treats revision as an exercise of judgment rather than a purely mechanical removal of mistakes. [1]

We can examine this practice diagnostically. A writer remembers what a passage was intended to say. That memory can supply a connection missing from the actual words. A later reading, or a reading by another person, changes the conditions of observation. The passage remains the same, but the reader is less able to complete it from the author's immediate intention.

Consider a hypothetical instruction: "After checking the value, it should be recorded." The writer may know who performs the check and where the record belongs. A reader entering the process for the first time does not. Merely replacing one word may leave the missing relationship untouched. Revision has to recover the activity and its participants before it can produce an adequate instruction.

A useful distinction follows between local and structural correction. A local alteration repairs an identifiable feature, such as a repeated word. A structural alteration changes the order or relationship of information. The two interact. Removing a paragraph can leave a pronoun without an antecedent; moving an explanation can make an earlier statement difficult to understand. A repaired passage must be reread in its new surroundings.

Quintilian's warning about excessive revision also has a diagnostic interpretation. A writer can continue changing a passage after the original problem has been resolved. Each alteration creates another version and another opportunity for loss. The investigation needs a stopping condition: the intended relationship is expressed, the identified ambiguity is resolved, and the remaining differences are matters of choice.

This is not evidence that ancient rhetoric anticipated software testing in a technical sense. It is evidence that writers already reflected on the conditions under which their own judgment could become more reliable. The history of writing diagnostics includes the deliberate alteration of those conditions, not only the invention of external instruments.

3. Manuscript correction and successive hands

The fourth-century Codex Sinaiticus preserves a text that was repeatedly examined and altered. Its modern project documents corrections associated with production as well as later interventions. The transcription distinguishes hands and classes of correction where the evidence permits, while retaining uncertainty where an attribution cannot be made. It also explains how the display represents readings that occupy margins or cross modern divisions of the text. [2, 3]

The manuscript gives us more than a corrected reading. It can preserve a relationship between an earlier inscription and a later intervention. Erasure, overwriting, an insertion, and a marginal addition are different kinds of evidence. Their location matters. A transcription that records only the final wording can hide the sequence through which that wording became available.

Let us consider a hypothetical omission repaired above a line. The added words may have been supplied by the original scribe, by someone checking against another exemplar, or by a later reader. The present appearance alone may not decide among these possibilities. Differences of ink and hand can contribute evidence, but they must be interpreted together with the page and its history. A confident reconstruction should not be manufactured where the evidence supports only alternatives.

Copying also creates dependencies among witnesses. If several manuscripts descend from one already altered copy, their agreement does not provide several independent confirmations of the reading. Conversely, a single surviving witness may preserve a reading lost elsewhere. Counting copies is therefore different from evaluating the relationships among them.

From our viewpoint, the manuscript is both a text-bearing object and a record of work performed on that text. These roles may conflict during presentation. A clean reading edition makes continuous reading easier. A diplomatic transcription tries to preserve selected features of one witness. An apparatus exposes alternatives and editorial decisions. Each makes something observable while placing other features outside its immediate view.

The modern digital representation adds another stage requiring examination. A correct transcription can be displayed misleadingly; a page image can be legible while the searchable text omits an addition. The old correction and its new representation need separate verification. The history of the artifact continues when it is digitized.

4. Printing traditions beyond a European starting point

A history of publishing cannot begin by treating European movable type as the origin of printing. The British Library’s Diamond Sutra scroll is dated 11 May 868 and was produced by woodblock printing. Its colophon records the sponsor, the purpose, and the date. UNESCO identifies the surviving second volume of Jikji, printed in Korea in 1377, as the oldest surviving evidence of printing with movable metal type. These are distinct technologies and distinct historical contexts. [4, 5]

The Tripitaka Koreana woodblocks provide a further example of a text embodied in a durable production resource. The surviving blocks and their preservation have a history different from that of any one impression made from them. UNESCO’s account treats the collection as documentary heritage and describes its place in Korean Buddhist printing. [6]

These examples help us separate the text, its production matrix, and its copies. A discrepancy in one impression may arise during printing. A discrepancy carved into a block can recur in impressions made

from that block. A damaged or replaced component can change the evidence available to later investigators. The location of a defect determines the scope of its possible propagation.

Consider, as a hypothetical comparison, an omitted character in a carved block and an omitted character in movable-type composition. Both can produce a missing character on the page. The operations needed to repair the production resource differ. The traces left by the repair may also differ. We should not infer one process merely from the similarity of the resulting page.

Publication records can identify a sponsor, place, or date without documenting every act of checking. A colophon is evidence for the statements it contains, interpreted in context. It is not a complete production log. Likewise, the survival of an accurately printed passage does not by itself prove a particular quality-control procedure. Claims about checking practices require their own evidence.

The comparative value lies in the questions we can carry across these cases. What was retained as the authority for the text? What material held the reproducible representation? Which intervention affected one copy and which affected future copies? Where did responsibility for a correction reside? The answers need to be established within each tradition rather than supplied by a universal story of progress.

5. Textual criticism and the choice of a reference text

Textual criticism examines relationships among surviving readings and the texts an edition seeks to present. A witness may preserve an omission, a substitution, an authorial revision, or a change introduced during transmission. The immediate task is to establish what the evidence supports. Deciding which reading to print is a further editorial act.

W. W. Greg's "The Rationale of Copy-Text," published in Studies in Bibliography in 1950–1951, made an influential distinction between substantive readings and the formal features often called accidentals. His argument did not give a selected copy-text unlimited authority over every reading. The choice of a base text and the judgment of particular variants remained distinguishable questions. [7]

For our purposes, the important diagnostic relationship is between observation and emendation. An observed reading belongs to a witness. An emendation belongs to an editor's reconstruction or correction. The latter may be well supported, but it should not be presented as if it were simply another visible inscription. A critical apparatus can preserve the distinction for later examination.

A hypothetical passage illustrates the problem. Two witnesses contain an awkward sentence; a third contains an elegant one. Elegance alone cannot establish priority. The third witness may preserve the author's revision, or it may contain a copyist's improvement. The awkward sentence may be damaged, or it may preserve a deliberate construction. We need evidence about the witnesses and the editorial aim before selecting a reading.

Modern encoding can make this structure explicit. The Text Encoding Initiative provides mechanisms for representing apparatus entries, readings, and their witnesses. Such representation enables alternatives

to remain available instead of disappearing behind the displayed text. The encoded distinctions are still editorial decisions and need documentation. [8]

There is consequently no single correction target for every edition. An edition of one historical document, a reconstruction of an earlier text, and an edition concerned with stages of authorial revision can legitimately present different results. Diagnostic rigor requires that the target be stated and that the route from evidence to presentation remain inspectable. Otherwise a disagreement about purpose can be mistaken for a failure of accuracy.

PART II

The printing house and the published record

6. Composition and correction in the printing house

Printing distributed the making of a text across several activities. Copy had to be interpreted and set; the resulting composition had to be checked; sheets had to be printed and assembled. The work created opportunities for error at the transitions as well as within each activity. A correct reading on a manuscript page did not automatically become a correct reading in every finished copy.

Joseph Moxon's Mechanick Exercises on printing, issued in 1683–1684, documents the work of compositors and correctors. His discussion recognizes a tension between following copy and repairing spelling or punctuation. The compositor's task therefore involved judgment even where fidelity to the supplied copy was the stated rule. [9]

Anthony Grafton's historical study The Culture of Correction in Renaissance Europe examines the people and practices involved in correcting printed books. It places correction within the social organization of scholarship and printing. This is a useful corrective to a history that attributes textual reliability to the press alone. [10]

Let us follow a hypothetical name through the printing house. The manuscript contains a difficult letter. The compositor selects a plausible reading. A corrector reads the proof against a second copy containing the same ambiguity. Both stages can agree and still produce the wrong name. The failure lies in a shared source of uncertainty, not necessarily in the absence of a second reader.

The evidence needed to investigate the case depends on what survives. Manuscript copy may localize the ambiguity. A marked proof may show that the name was queried. A later letter may record the intended spelling. If only the published book remains, the range of defensible explanations is wider. The lack of a surviving query cannot establish that no query was made.

A printing house can thus be examined as a network of transformations and responsibilities. The comparison with a software pipeline is useful at this level. Both require attention to inputs, intermediate states, and handoffs. The historical mechanisms remain different, and a successful explanation must identify the actual operation that produced the discrepancy.

7. Proofs and the visibility of production work

A proof makes an intermediate result available for examination. It is especially valuable because it exposes consequences that cannot be judged from the manuscript alone. The words may be correct while the line arrangement, page furniture, or relationship between a caption and an illustration is

wrong. Proofreading therefore has a different object from composition, even when the same person performs both tasks.

Moxon's account of printing provides a historical description of proof correction within the production process. Modern professional guidance retains the distinction between work on copy and checking a text prepared for publication. The scope varies with the commission, but the stage and authority of the proof remain central. [9, 11]

The mark made on a proof is itself a message. It identifies a location, requests an operation, and may indicate the replacement. An ambiguous correction can become the cause of a new error. A crossed-out word without a clear insertion point, for example, can leave the compositor uncertain whether the instruction is deletion, substitution, or movement. The communication of the repair needs verification alongside the repair itself.

Consider a hypothetical page on which a figure moves after a paragraph is shortened. The figure is now separated from the passage that explains it. The original textual correction is valid, but its production consequences are undesirable. Reading only the amended sentence will not detect the problem. We need to inspect the page and the neighboring page after the alteration has been applied.

This gives us two distinct comparisons. The first checks the proof against approved content. The second checks the corrected proof against the requested changes. The latter asks whether the repair was implemented as intended and whether it introduced another discrepancy. A correction list that is marked complete without this second comparison records an administrative state rather than demonstrated correctness.

The proof also has limits. Its authority depends on its relationship to the actual production state. An earlier proof can remain visually convincing after the source has changed. A proof made for one output process may not predict another. Its identifying information and production conditions are therefore part of its diagnostic value.

8. Errata cancels and the persistence of an error

An error discovered after printing cannot always be removed from every copy. Publishers can provide a correction list, insert a slip, replace a leaf, or correct a later impression. These interventions affect different populations of copies. Bibliographical investigation consequently asks which state a particular copy preserves, rather than assuming that all copies bearing the same title are identical. [10, 12]

The 1631 Bible commonly called the Wicked Bible provides a familiar example of a small textual omission with a large semantic effect. The Folger's exhibition material records the misprinted reading "Thou shalt commit adultery" in Exodus 20:14. The example is useful without embellishing it with an unsupported explanation of how the omission occurred. The surviving reading establishes the discrepancy; it does not, by itself, identify the complete production mechanism. [13]

We can distinguish detection, correction, and dissemination of the correction. Detection establishes that something is wrong. Correction produces an amended representation. Dissemination makes the amendment available to the people using the earlier representation. Success at the second stage does not guarantee success at the third. A reader may continue using a copy that never receives the inserted slip.

A hypothetical technical handbook makes the distinction practical. An erratum corrects the unit attached to a value. A later printing incorporates the correction. A training sheet copied from the first printing retains the old unit. The book has been repaired, but the dependent document has not. The diagnostic object has expanded from a page to a family of publications.

Errata are also historical evidence. They identify discrepancies recognized by a particular time and can reveal how the producer understood responsibility. They are not necessarily exhaustive, and their silence about a passage does not certify it. Nor should an erratum be accepted without checking that it belongs to the edition and state under examination.

The persistence of old copies becomes especially visible in digital publishing, but the underlying problem predates it. Publication distributes evidence and distributes errors. An adequate correction must identify its target precisely enough that readers can determine whether it applies to the copy they hold.

9. Publication control and the authority to alter

The history of correction includes interventions that do not aim at textual accuracy. A passage can be removed because it is politically unwelcome, institutionally inconvenient, commercially risky, or inconsistent with an editor’s preferred presentation. The altered text may be cleaner or more acceptable to one authority while preserving less of the author’s intended argument.

Milton’s Areopagitica, published in 1644, argues against prior licensing of print. It belongs to a particular religious and political setting and should not be treated as a statement of every modern position on freedom of publication. Its relevance here is the conflict it exposes between the production of a text and the authority that permits it to circulate. [14]

For diagnostic purposes, an alteration has both a textual effect and an institutional explanation. A deleted qualification changes a claim. The deletion might result from authorial revision, editorial shortening, a licensing demand, or accidental loss. The visible difference does not identify its motive. Correspondence, instructions, contracts, and successive states may be needed to reconstruct the intervention.

Consider a hypothetical reviewer who requests that a cautious conclusion be made more decisive. If the author accepts the request, the final sentence is authorized in one sense. It may nevertheless be less faithful to the available evidence. Authorization and evidential adequacy are separate properties. A production record should make it possible to examine both.

The same distinction applies to contemporary style normalization. A house convention can legitimately govern spelling or reference form. It does not automatically justify replacing an author's technical distinction with a more familiar phrase. The editorial brief needs to identify the scope of permitted change and the questions that require consultation.

We therefore avoid equating every intervention with debugging. Some alterations repair defects; some create a new version; some suppress material; some change the conditions under which readers can interpret it. Historical diagnosis begins by identifying the kind of intervention before assessing its success. Calling every change a correction would erase precisely the differences we need to investigate.

10. Material bibliography and the comparison of copies

Printed copies can preserve evidence of their production in small differences of type and text. At the Folger, Charlton Hinman used an optical collator to compare copies of Shakespeare's First Folio. By alternating aligned views, the instrument made differences conspicuous. His examination of 55 copies contributed to The Printing and Proof-reading of the First Folio of Shakespeare, published in 1963. [12]

The instrument changed the conditions of comparison. A reader following the meaning of a sentence can pass over a small typographic difference. An optical comparison draws attention to the visual discrepancy itself. The observation is then available for interpretation in relation to printing practice, correction, and the history of individual copies.

This separation between detecting a difference and explaining it remains important. A changed letter may indicate a textual correction, damaged type, or another production event. The collator does not supply that explanation automatically. It makes a candidate discrepancy visible so that bibliographical reasoning can proceed.

A hypothetical pair of copies illustrates why a single representative copy can be insufficient. One preserves an uncorrected sheet and another a corrected sheet. If only one is examined, a description of the edition may generalize from a local state. Comparing further copies can reveal the distribution, although the surviving sample may still be incomplete.

Material bibliography also provides a useful limit to purely textual comparison. Two pages can yield the same extracted words while differing in typography, impression, or arrangement. Conversely, digitization may introduce a difference into the transcription where the printed pages agree. The investigation should retain the representation in which the discrepancy was observed.

We can now see several forms of diagnostic evidence operating together: the reading, the physical impression, the relationship among copies, and the documented production practice. None substitutes for all the others. A convincing explanation connects them without treating a reconstructed sequence as if it were a directly recorded log of the printing house.

PART III

Composition style and prose debugging

11. Drafts and the reconstruction of writing

A published text shows one selected state of a work. Drafts can expose alternatives, abandoned passages, and changes in organization. They allow us to ask how the work was made, while also showing why its history cannot be reduced to a list of corrected mistakes. A deleted passage may have been successful in itself and still have become unnecessary within the developing whole.

The manuscripts of Madame Bovary provide a substantial example. Rouen's library describes their digitization in 2002–2003 and the opening of the associated website in 2009, in partnership with the university's Centre Flaubert. The Centre Flaubert supplied the transcription and genetic table used by the project. The underlying manuscripts belong to the nineteenth-century writing process; their online arrangement is a later scholarly representation. [15]

The distinction matters because the sequence of surviving leaves is not automatically the sequence of composition. A page may have been reused, an insertion may have been written separately, and a passage may have moved. A genetic reconstruction proposes relationships among such evidence. It should retain the grounds for those relationships and the uncertainty that remains.

Let us consider a hypothetical paragraph present in three drafts. The first supplies a detailed explanation, the second removes it, and the third adds one sentence elsewhere. A simple count of additions and deletions would miss the functional change. The explanation has been redistributed. To understand the revision, we need to follow the idea across locations, not merely compare the same numbered paragraph.

This suggests a useful viewpoint for contemporary authoring. Stable identifiers for passages, terms, and claims can help us follow a developing work even when page numbers change. The historical manuscript does not provide those identifiers automatically. An editor may supply them as part of a documented analytical model.

Drafts also preserve unrealized possibilities. Calling every earlier state defective would turn the final publication into the inevitable destination of the process. A diagnostic history should allow for branching, reconsideration, and changes of purpose. The evidence can show that a writer chose differently without establishing that every rejected choice was an error.

12. Revision as a recursive activity

Twentieth-century research on composition examined revision as part of writing rather than merely its final cleanup. Nancy Sommers's 1980 study compared revision strategies of student and experienced

writers. Linda Flower and John R. Hayes's 1981 cognitive process theory described interacting processes of planning, translating ideas into text, and reviewing, organized around goals that can themselves develop. These studies challenged a simple linear sequence of planning, writing, and then correcting. [16, 17]

The diagnostic consequence is that a difficulty in a sentence may originate in an unresolved problem of thought. The writer may not yet know which distinction the paragraph needs to make. Repeatedly replacing words can leave that problem untouched. A change in the argument may be required before a satisfactory sentence becomes possible.

Consider a hypothetical account that alternates between "observation" and "measurement." The variation might be stylistic, or it might conceal two different operations. The editor should first determine what the author intends each term to denote. Once that relationship is clear, consistency can be checked. Before it is clear, uniform replacement risks destroying a distinction still being developed.

Revision can also change the specification against which the text is judged. A chapter initially written for specialists may later be intended for a broader readership. Passages that were adequate under the original assumption may now require explanation. The resulting work is not simply an error-corrected version of the old chapter; it has a changed communication problem.

We can preserve these differences in revision records. A correction repairs a recognized discrepancy within the existing purpose. A restructuring changes how the purpose is achieved. A change of scope alters what the text is meant to do. These categories are analytical aids, and one intervention can occupy more than one category. Their value lies in making the reason for the change available.

The process viewpoint also helps explain why a final reading is necessary. A sequence of locally reasonable revisions can produce an incoherent whole. The text needs to be examined again as a continuous argument, because its present relationships were not all visible when the individual changes were made.

13. Style manuals and shared conventions

A style manual provides a shared reference for recurring editorial decisions. The University of Chicago Press traces its manual from internal guidance to the first book published in 1906. William Strunk's The Elements of Style, first published in 1918, offered a compact set of principles for usage and composition; the early text should be distinguished from later versions associated with E. B. White. [18, 19]

These works serve different purposes, but both make selected judgments reusable. An editor need not reopen every question about capitalization, punctuation, or presentation from the beginning. A convention can be recorded, applied, and checked across the document. This reduces variation caused by independent local decisions.

The existence of a rule does not establish that it applies everywhere. A quotation may need to preserve its source. A title may use an official spelling that differs from house style. A character's speech may

deliberately violate an expository convention. The style sheet should record such cases as intended exceptions rather than allowing them to appear as unexplained inconsistencies.

Consider a hypothetical book that uses “trace” for captured events and “narrative” for an interpretation of those events. An editor seeking variety replaces several occurrences of “trace” with “narrative.” The sentences remain fluent, but the terminology no longer supports the argument. The relevant style rule is the preservation of a technical distinction, not the avoidance of repetition.

A style sheet can therefore function as a limited specification. It records decisions about names, terms, spelling, numbering, and other repeated features. It cannot fully specify the author’s voice, nor can it guarantee the truth of a claim. Its diagnostic value depends on recognizing that boundary.

There is also a temporal dimension. A convention may change between versions of a manual or between editions of a publication. A later editor needs to know which rule governed the earlier work. Otherwise an intentional historical usage can be misclassified as carelessness. Shared conventions improve consistency when their scope and date accompany them.

14. Readability measures and their limits

Readability research made selected textual properties available for numerical comparison. Rudolf Flesch’s 1948 paper “A New Readability Yardstick” is a major historical example. Its reading-ease approach uses sentence and word-length characteristics as indicators of difficulty. Such measures turn aspects of prose into inspectable quantities, but they do not directly measure the truth, completeness, or explanatory value of an argument. [20]

Let us consider two hypothetical instructions. “Close the valve” is short and easy to parse. It may still be unusable if several valves are present and none is identified. A longer instruction can be more effective because it specifies the correct component and the required condition. A reduction in sentence length does not necessarily constitute a repair.

Metrics are most useful when the measured property is connected to the communication problem. Unusually long sentences can be candidates for review. They may contain excessive embedding, but they may also express a relationship that becomes harder to follow when divided. The number identifies a place to inspect; the diagnosis requires reading.

A document can also improve its score while losing its subject. Replacing an established technical term with a shorter everyday word may lower apparent difficulty and create ambiguity. Removing qualifications can simplify syntax while making a claim inaccurate. These are examples of optimizing an indicator at the expense of the intended communication.

The same problem appears when several kinds of judgment are compressed into one quality score. Spelling, grammar, structure, source support, and suitability for a reader are related but distinct. A strong result in one area cannot compensate automatically for a serious failure in another. A perfectly punctuated false statement remains false.

We therefore treat readability measures as one viewpoint on a text. They can support comparison between versions or draw attention to outliers. They should be accompanied by the population, language, genre, and purpose for which the interpretation is being made. The historical contribution is the availability of a repeatable observation, not the replacement of editorial judgment by a number.

15. BUGS in Writing and the explicit debugging metaphor

Lyn Dupré's BUGS in Writing: A Guide to Debugging Your Prose appeared in 1995. Addison-Wesley published BUGS in Writing, Revised Edition: A Guide to Debugging Your Prose as the second edition in 1998, ISBN 978-0-201-37921-1. The revised edition is the book considered here. [21, 22]

Dupré presents recurring writing problems through short segments and classified examples. The categories bad, ugly, good, and splendid distinguish technical error from less satisfactory expression and from improvement within acceptable writing. Her approach also encourages writers to recognize the particular problems that recur in their own prose. This makes the debugging metaphor explicit while leaving room for judgment and the development of an ear for language. [22]

We can connect that approach with our own pattern-oriented diagnostics. A recurring problem becomes easier to recognize when its characteristic form has a name and examples. Recognition, however, should lead to examination of the present context. A familiar surface form may arise for different reasons, and the same local alteration will not repair every instance.

Consider our hypothetical sentence: "The editor told the author that she had removed the qualification." The sentence is grammatical, but the referent of "she" may be unclear. Replacing it with "the editor" resolves the ambiguity only if the editor actually performed the removal. A mechanically clearer sentence can become factually wrong. The repair requires knowledge of the event being described.

This example also separates diagnosis from optimization. Resolving the referent repairs an identified communication problem. Choosing a more graceful sentence afterward may improve the prose without repairing a further defect. The two activities can be connected while retaining different stopping conditions.

For a history of writing and publishing, Dupré's book is therefore a documented meeting point between computing vocabulary and editorial practice. It does not establish that all writing problems are software bugs. It invites a disciplined examination of recurrent prose difficulties, and we can extend that examination to the artifacts and transformations through which a text becomes a publication.

PART IV

Editorial work and the verification of claims

16. Structural editing copyediting and proofreading

Different editorial activities examine different objects. Structural editing asks whether the work contains and arranges what its purpose requires. Line editing examines the expression and movement of the prose. Copyediting commonly attends to correctness and consistency within the editorial brief. Proofreading examines material prepared for publication. Professional boundaries vary, and the Chartered Institute of Editing and Proofreading emphasizes the need to distinguish the services and agree their scope. [11]

We can understand the distinction through the evidence available at each stage. A structural editor may need the entire argument and the intended readership. A copyeditor may need a style sheet and authoritative forms of names. A proofreader needs the actual pages and, where applicable, the approved copy or preceding proof. Giving each person the same file without the relevant context does not create equivalent opportunities for diagnosis.

A hypothetical missing explanation can pass through several competent checks. The copyeditor sees grammatical sentences. The proofreader sees clean pages. Neither was asked to assess whether the chapter teaches the concept it promises to teach. The omission persists because the required question was not assigned, rather than because every participant failed at the same task.

Conversely, late structural intervention can disturb completed production work. Adding a paragraph may change page breaks, illustration positions, cross-references, and index locators. The change may be necessary, but it reopens dependencies. Calling it a small textual insertion understates the scope of verification it requires.

A useful editorial record connects each query with its disposition. The record should show what was questioned, what evidence or decision resolved it, and which version incorporates the result. A query that disappears from the visible document is not necessarily a resolved query. It may have been deleted accidentally or lost during exchange.

The history of professional editing can thus be read as a history of differentiated attention. Reliability depends on complementary examinations and clear handoffs. The number of passes matters less than whether those passes examine the relationships that can actually fail.

17. Terminology navigation and document structure

A book is more than a sequence of sentences. Headings, captions, notes, references, contents entries, and index locators connect distant parts. These relationships can fail while each local passage remains readable. Their examination requires a view of the document as a connected structure.

Editorial conventions provide ways to manage such relationships, but the underlying choices must first be made. A term needs a stable meaning. A heading needs a defined scope. A figure reference needs a target. Once these relationships exist, numbering and presentation can be checked against them. Consistency of appearance alone is not enough.

Let us consider a hypothetical glossary entry for “proof.” In a publishing history, the word can denote a trial impression or a demonstration of a proposition. If both meanings occur, a single undifferentiated entry may mislead. The index may need separate subentries, and the local passage may need clarification. The problem concerns semantic organization before it concerns typography.

Generated navigation creates another distinction between a relationship and its displayed value. A cross-reference can point to the correct heading while showing an earlier page number. A manually typed reference can show the right number today and become wrong after pagination changes. The diagnostic question is which mechanism maintains the connection and when its result was last regenerated.

An index introduces selective judgment. Its entries should lead readers to substantive discussion, not simply to every occurrence of a word. A page can contain a term only in a passing comparison. Another page may discuss the concept without using the preferred term. Mechanical search supports index preparation but does not settle its intellectual organization.

The practical consequence is that navigation needs its own verification after the document stabilizes. Contents, cross-references, captions, notes, and index locators should be checked in the final representation. This is a proposed diagnostic discipline drawn from the relationships examined here: inspect the connections at the stage where readers will use them, and preserve enough information to rebuild them after a change.

18. Fact checking references and the evidence behind a sentence

A citation performs several functions that need separate checks. It identifies a source, directs the reader to it, and implies a relationship between that source and the nearby claim. A correct title and working link establish only part of that relationship. The source may exist without supporting the statement for which it is cited.

Magazine publishing also developed specialized responsibility for checking claims. The New Yorker’s historical account of its fact-checking department describes the institution’s own development and practices. Such departmental history reminds us that factual verification is organized labor with a remit,

records, and constraints. It should not be treated as an automatic consequence of good copyediting. [23]

Digital identifiers and bibliographic metadata make source discovery and matching more systematic. Crossref provides metadata retrieval and reference-matching facilities, while its linking model connects identifiers with registered works. These facilities support identification. They do not determine whether a quotation is accurate or whether a particular inference follows from the cited passage. [24]

Consider a hypothetical sentence claiming that a method was invented in 1983. Its source is a paper published in that year that describes an already existing method. The reference is real, the date is correct, and the historical claim is still unsupported. The paper establishes documentation by that date; priority requires additional evidence.

The same issue arises with editions. A later edition may add a passage absent from the original. A digital repository may assign a recent posting date to an older work. A catalog may describe a reprint. The investigation should separate the event, the original publication, the consulted edition, and the present access route. These dates answer different questions.

Fact checking also needs to retain qualifications. A source reporting an association does not automatically support a causal claim. A result obtained for one sample does not establish universal behavior. A sentence can drift during editing when a cautious verb is replaced with a decisive one or when the condition limiting a result is removed.

We can make the evidence relationship inspectable by retaining a claim with its source location and a short statement of what that source supports. This is especially useful for claims of priority, numerical results, quotations, and disputed attributions. The record does not replace reading. It makes the next reading more focused and helps prevent a correct reference from acquiring an incorrect role in the argument.

19. Peer review and the limits of an editorial judgment

The Royal Society’s Philosophical Transactions began in 1665 under Henry Oldenburg. Its history distinguishes that origin from the more systematic expert refereeing introduced in the nineteenth century; the Society identifies 1832 as a significant point in that development. It would therefore be misleading to project the full apparatus of contemporary peer review back onto the journal’s first issue. [25]

Peer review makes a submitted argument available to readers with relevant expertise. Reports can expose unsupported conclusions, missing comparisons, unclear methods, and disagreements about interpretation. The reports are also artifacts of a particular review, shaped by the questions asked and the material supplied. Acceptance is an editorial outcome, not a certificate that every statement has been independently verified.

Let us consider a hypothetical reviewer who checks a mathematical derivation but does not follow every bibliographic link. Another reviewer checks the empirical method but not the software that produced a figure. Their combined work may be valuable while leaving a reference error or rendering defect undiscovered. The scope of the evidence should remain visible when the outcome is described.

NISO's 2008 recommendations for journal article versions distinguish states including submitted manuscript under review, accepted manuscript, proof, and version of record. These terms make it possible to identify which artifact a review or correction concerns. An accepted manuscript and a published article are related states, but production can introduce differences between them. [26]

Disagreement among reviewers is not automatically evidence that one has failed. They may identify different risks, apply different standards, or read an ambiguity in different ways. A useful editorial response specifies which concern is being addressed and why the resulting change is appropriate. Merely accumulating revisions can produce a text that satisfies individual comments while losing coherence.

The diagnostic value of review lies in the additional viewpoints and the reasoning they preserve. To use it well, we need to connect the report, the response, and the revised text. Otherwise the publication records the outcome but loses the path by which a problem was examined.

20. Translation and the conditions of a correct representation

Translation introduces another relationship requiring diagnosis: the relationship between a source passage and its expression in another language. A fluent target text can omit a qualification, reverse a relation, or make a technical term less precise. A literal rendering can preserve words while failing to communicate the intended relationship. The criterion of correctness therefore depends on both meaning and purpose.

The published representation adds further conditions. Unicode normalization defines equivalences among certain character sequences, while the handling of scripts such as Arabic includes directionality and shaping requirements. W3C's work on Arabic and Persian layout documents why text processing cannot assume the conventions of left-to-right Latin-script prose. [27, 28]

Consider a hypothetical bilingual paragraph containing an Arabic sentence, a Latin-script product name, and a number in parentheses. The stored characters may be correct while the visual order becomes misleading. Rewriting the translation would address the wrong layer. The relevant evidence includes the character sequence, directional markup, font, and rendered result.

A different failure can occur when two visually similar forms are treated as identical by an editor but remain distinct in search or indexing. Normalization can help with defined equivalences; it should not be confused with indiscriminate replacement of characters. Scholarly transcription, names, and quoted material may require distinctions that ordinary office workflows overlook.

We can separate several examinations. A language reviewer checks the meaning and idiom. A subject specialist checks terminology. A production review checks script behavior and layout. An accessibility review checks whether the document exposes the intended language and reading order. A successful result in one examination does not establish the others.

The broader historical lesson is that publishing technologies carry assumptions about writing. When a system moves into another linguistic context, those assumptions become visible through failures. Diagnosis should identify the assumption that no longer holds rather than treating the language itself as an exception to a supposedly universal document model.

PART V

Computers and the production of documents

21. Mechanization and editable text

Mechanized writing changed the cost and visibility of revision. A typewritten page could be corrected or retyped, but the clean copy often concealed the sequence of decisions behind it. Electronic storage made it possible to alter a retained representation and produce another output without recreating the entire page manually. This changed both the opportunities for correction and the evidence that might survive.

Mechanization also transformed composition for print. The Science Museum Group dates Tolbert Lanston's Monotype invention to 1887 and preserves material relating to its subsequent development. Its collections include a Lumitype 540 photocomposer from around 1965. These examples mark distinct technical arrangements for composing type, including mechanical casting and photographic production. [29, 30]

Each arrangement changes the likely mechanisms and locations of a fault. A wrong character can originate in input, selection, a production component, or later handling. Increased output capacity does not eliminate checking; it changes the amount of material and the technical dependencies that checking must address.

IBM introduced the Selectric typewriter in 1961 and added magnetic-tape storage in the MT/ST in 1964. The latter stored characters on magnetic tape as part of document preparation. At Xerox PARC, work on interactive systems included the Bravo editor, designed by Butler Lampson and Charles Simonyi and implemented by Simonyi and colleagues. These developments belong to a broader history of word processing rather than a single uncontested invention event. [31, 32]

Stored text separates the editable representation from any one printout. That separation is productive, but it creates a new diagnostic question: which state produced this page? A file may have changed since the page was printed. A clean copy may have been generated before the last correction. The most recently opened file is not necessarily the source of the distributed document.

Let us consider a hypothetical author who corrects a name on screen and then prints from a second open copy. The correction exists, but the output does not contain it. Reading the corrected source alone cannot explain the reader's complaint. We need the actual output and the relationship between the two copies.

Electronic editing also changes what deletion means. A removed passage may survive in a backup or revision record, or it may disappear from all retained states. An undo history is not the same as a durable archive. The capacity to reverse an action during a session does not guarantee that another investigator can reconstruct it later.

The diagnostic opportunity introduced by editable state is therefore conditional. When states are identified and retained, comparison becomes easier. When they are repeatedly overwritten without a record, a polished final document can conceal more of its history than a heavily marked paper draft. The technology enables observability; the workflow determines how much of it is realized.

22. Spelling grammar and the Writer's Workbench

Computational writing aids made repeated textual examinations practical at larger scale. Douglas McIlroy's 1982 "Development of a Spelling List" describes the construction and use of a word list for spelling checking. The problem included both vocabulary selection and economical representation. A spelling checker depended on an engineered language resource, not on unrestricted knowledge of what a writer might intend. [33]

The UNIX Writer's Workbench extended assistance beyond spelling. Nina H. Macdonald's 1983 account describes programs for proofreading, stylistic analysis, alternative presentations of text, and reference information. The system treated evaluation and editing as activities that could be supported by several specialized tools. [34]

This division remains diagnostically useful. A dictionary comparison can identify an unfamiliar token, while a style check can identify a construction worth reviewing. Neither observation alone establishes an error. A technical name may be absent from the dictionary. A long passive construction may be appropriate where the actor is unknown or irrelevant.

Consider the hypothetical sentence "The form was sent to the manger." Every word may be present in an ordinary dictionary. The error concerns the intended word in context. Conversely, the name of a new product can be correctly written and still attract a warning. False negatives and false positives arise from the relationship between the tool's model and the document's vocabulary.

Automated replacement introduces a further hazard. A global change can repair one use of a term and corrupt a quotation, a file path, or a different technical expression. The change set needs examination. A tool that reports no remaining matches has demonstrated only the absence of the searched form under its matching rules.

We can view these systems as an early repertoire of computational diagnostic instruments for prose. Their strength lies in repeatable attention to selected features. Their limitation is also selected attention. A sound workflow uses the reports to direct examination and records accepted exceptions where recurrence would otherwise generate noise.

23. Typesetting programs and the separation of source from output

Computerized typesetting made the document-production mechanism an explicit computational object. Knuth's TeX work and its published documentation connected typography with algorithms and

reproducible source. Knuth and Michael Plass's 1981 paper on paragraph breaking examined the paragraph as a whole instead of choosing each line break in isolation. Layout could now be investigated through the rules and optimization criteria that generated it. [35, 36]

Desktop publishing offered another mode of interaction with the page. Adobe's account of the development of PDF places its launch with Acrobat in June 1993, following the role of PostScript in desktop publishing. These technologies made it possible to exchange a representation of page appearance across systems, while introducing their own rendering and production dependencies. [37]

The distinction between source and output is central. The source may contain the correct text and an incorrect instruction. It may contain a correct instruction interpreted differently by another environment. The output may be valid as a file and still present a poor line break or a missing glyph. Each condition requires evidence from a different layer.

A hypothetical mathematical book illustrates the point without requiring a formula. The source contains a symbol that displays correctly on the author's computer. A substituted font in another environment lacks the glyph. The apparent gap on the page is not a missing word in the manuscript. Editing the sentence to conceal the gap would leave the production dependency unresolved.

Typesetting diagnostics therefore need both semantic and visual examinations. Text extraction can detect missing passages or duplicated content. Rendered pages reveal clipping, collisions, page breaks, and visual relationships. A clean text comparison does not certify the page, and a clean page does not certify the underlying reading order.

Once the production mechanism is software, some publishing failures become literal debugging problems. A minimal example, a recorded tool version, and a reproducible sequence can help localize the defect. Other failures remain authoring or configuration errors. The first task is to determine which kind of explanation fits the observed evidence.

24. Structured text and machine checkable relationships

Markup makes selected relationships in a document explicit. A heading can be represented as a heading rather than merely as large type. A citation can be associated with a bibliographic record. A note can be linked to its anchor. These distinctions allow software to inspect structure and produce several presentations from related source material.

The Text Encoding Initiative released its P3 Guidelines in 1994. Its later guidelines include detailed support for textual witnesses and critical apparatus. In scholarly journal publishing, NISO published the JATS standard in 2012, defining elements and attributes for article content and metadata. The two initiatives have different histories and scopes, although both make document structure available for systematic processing. [8, 38, 39]

Validation establishes conformance to specified structural rules. It can detect a missing required element, an invalid arrangement, or an unresolved identifier where the rules provide for that check. It

cannot automatically establish that a paragraph's historical claim is true or that the chosen heading accurately describes its content. The structure and the interpretation need different examinations.

Let us consider a hypothetical reference list in which every entry has the required fields. One entry names the wrong article. The document validates because the fields are structurally acceptable. Bibliographic verification is still required. In another case, the entry is correct but its identifier is duplicated. The text can look satisfactory while the structured relation is defective.

Single-source publication also creates shared dependencies. A corrected paragraph can flow to several outputs, reducing repeated manual work. An incorrect transformation can propagate to all of them. When an error appears everywhere, agreement among outputs may indicate shared ancestry rather than independent confirmation.

Structured text is therefore a means of increasing observability and controllability over selected document relationships. Its benefit depends on the distinctions encoded and the checks actually performed. A schema that represents only appearance cannot verify a semantic relationship that was never recorded.

25. Preflight color proofs and the physical publication

Digital preflight examines properties that matter to a production process. Adobe's documentation describes profile-based document checks and inventory reports for resources such as fonts and color spaces. A preflight result is meaningful in relation to the chosen profile. A file can pass one set of checks while failing requirements that were not included. [40]

Physical proofing adds a different comparison. ISO 12647-7:2016 concerns hard-copy digital proofs intended to simulate a specified printing condition. The target is a defined production condition, rather than an unspecified resemblance between a monitor and a printed sheet. [41]

Consider a hypothetical cover that appears satisfactory on screen but loses fine text near the trim. The words and colors can be correct while the geometry is unsuitable for the intended manufacturing process. A separate problem occurs when an illustration's distinctions depend on colors that become difficult to separate in the chosen print process. The evidence must include the actual dimensions and production assumptions.

A proof is consequently a model of an intended result. Its usefulness depends on the features it represents. A screen preview may help assess composition and page sequence. A physical proof can expose paper, binding, trimming, and reproduction effects. Neither should be described as a complete simulation of every condition under which the publication will be read.

The diagnostic distinction between file, rendering, and manufactured copy is especially important for print-on-demand and small publishing operations. The same person may act as author, editor, designer, and production coordinator. The responsibilities remain distinct even when they are concentrated in one person.

A useful release examination therefore follows the publication through its actual transformations. It checks the source, the generated file, the rendered pages, and the relevant physical sample. The aim is to identify which representation supports each conclusion. A successful export establishes that a program produced a file; it does not, by itself, establish that a reader will receive a usable book.

PART VI

Versions distribution and preservation

26. Revision histories and the identity of a version

Electronic collaboration makes several kinds of change record available. Word can display tracked revisions and compare document versions. Git records selected repository states and provides comparisons among them. These mechanisms preserve different kinds of information and depend on how participants use them. A revision display, a stored version, and an approval record are related but separate artifacts. [42, 43]

Let us consider three copies called “final.” One contains the author’s corrections, one contains a copyeditor’s later changes, and one generated the distributed file. The filename is insufficient to identify their relationship. Timestamps help only when their meaning is understood; copying a file can change a timestamp without changing the text.

A content hash can distinguish byte-identical files from files that differ. It does not explain the difference or determine which file is approved. A revision identifier can connect a build with a source state, but only if the build record actually preserves that connection. The diagnostic value lies in the relationship recorded, not in the mere presence of an identifier.

Tracked changes also require careful interpretation. Hiding revision marks is not equivalent to accepting or rejecting them. Comparing two endpoints reveals their differences but cannot recover every intermediate decision. A comment can explain why a change was proposed without proving that the resulting text implements it correctly.

Plain-text workflows make many comparisons convenient, particularly for documentation maintained beside software. Rich documents can preserve other information more directly, including page layout and editorial annotations. The choice of representation should follow the work’s requirements. Version control is not limited to one tool, and no tool automatically records all relevant context.

The historical change is the increasing possibility of retaining a navigable sequence of states. To make that sequence diagnostically useful, we need to identify baselines, record the reasons for consequential changes, and connect outputs with their sources. Otherwise a large revision history can remain as difficult to interpret as a box of undated paper drafts.

27. OCR and the difference between an image and its text

Digitization can produce a page image and a machine-readable text layer. These are separate representations. Optical character recognition estimates text from visual evidence; it does not recover an infallible original. Ray Smith’s 2007 account of the Tesseract OCR engine describes a computational

approach to this task. The Library of Congress explicitly notes that OCR of historical newspapers can contain errors associated with markings on the original page, unusual text styles, and very small fonts. [44, 45]

The distinction changes how we interpret search. A word absent from an OCR search result may still be visible on the page. A phrase returned by search may be a recognition error. Negative search evidence is therefore conditional on the coverage and quality of the searchable representation.

Consider a hypothetical newspaper column in which a damaged letter causes a surname to be recognized as a common word. The image preserves enough evidence for a reader to identify the name, but the search index does not. A historical analysis based only on token counts may underestimate the person's appearance in the archive. The defect belongs to the derived data, not necessarily to the newspaper.

Layout recognition introduces further problems. A system can read two columns in the wrong order, attach a caption to the wrong image, or insert a running head into a paragraph. The individual words may be accurate while their sequence is wrong. Character accuracy alone cannot describe this failure.

Corrections to OCR also need provenance. A human-supplied reading may be certain, tentative, or reconstructed from context. Keeping the image and a record of the intervention allows another reader to evaluate it. Silently replacing an uncertain character with a plausible one can turn interpretation into apparently direct evidence.

The diagnostic workflow therefore compares representations according to the claim at issue. Search supports discovery. The page image supports examination of the visible inscription. The physical object may remain necessary where the image is inadequate. Each stage adds access while introducing conditions that historical interpretation must retain.

28. Electronic books accessibility and multiple reading conditions

An electronic publication is experienced through reading software and user settings as well as through its stored content. Reflow, enlarged text, alternate fonts, navigation, and assistive technologies can expose relationships that are invisible in one default view. The publication needs to work across the reading conditions it claims to support.

EPUB 3.3 first became a W3C Recommendation in May 2023. EPUBCheck checks conformance of EPUB publications, while EPUB Accessibility 1.1 addresses accessibility requirements and discoverability metadata. W3C's accessibility guidance also emphasizes that automated tools alone cannot establish accessibility. These are complementary examinations with different scopes. [46, 47, 48, 49]

Let us consider a hypothetical textbook whose chapter titles are visually prominent but represented as ordinary paragraphs. A sighted reader can recognize them on screen. A navigation system relying on semantic headings may not. The visual hierarchy and the machine-readable hierarchy disagree.

A second example concerns an illustration with an alternative description that says only “diagram.” The presence of an alternative-text attribute may satisfy a simple mechanical check. It does not give a reader the information needed to understand the argument. The relevant criterion is whether the description conveys the function of the illustration in its context.

Multiple output formats create related but distinct obligations. A printed page needs stable locators and legible physical presentation. A reflowable book needs resilient structure. An audio rendering needs a comprehensible reading sequence. The same source can support these outputs, but no single visual inspection establishes their equivalence.

Accessibility is therefore part of publication diagnostics, not a decorative addition after production. It asks whether the intended information remains available under different modes of reading. Failures can reveal hidden assumptions in the document model, such as the belief that proximity on a page always communicates association. Correcting those assumptions can improve the publication for many readers while preserving their different needs.

29. Metadata identifiers and distributed copies

Publishing distributes descriptions as well as content. Titles, author names, formats, identifiers, edition statements, and links travel through catalogs, booksellers, libraries, repositories, and reading systems. A correct file can become difficult to find or incorrectly described if these records disagree.

The International ISBN Agency distinguishes publications by relevant edition and format characteristics. Crossref’s DOI infrastructure connects registered identifiers with metadata and locations for scholarly works. These systems solve identification and linking problems within their respective scopes. An identifier does not certify the accuracy of the text it identifies. [24, 50]

Consider a hypothetical revised book whose cover and title page say Version 2 while a retailer’s description still refers to Version 1. A purchaser cannot easily determine which content is being supplied. The defect concerns the relationship among product description, file, and edition, even if each record is syntactically valid.

Distribution can also create divergence. One repository retains an accepted manuscript, another provides the version of record, and a personal page links to an earlier draft. The versions may all be legitimate, but their status must be explicit. NISO’s vocabulary for article versions helps distinguish such states. [26]

Broken links are only one class of failure. A link can resolve successfully to the wrong edition, a replacement landing page, or content whose status has changed. A successful network response establishes availability at that moment. It does not establish identity or suitability as evidence.

From a diagnostic viewpoint, the publication is a set of related objects distributed across systems. The investigation must follow the particular claim or discrepancy through those relationships. Correcting the

master record is necessary where that record is wrong, but the result should also be checked at the places where readers encounter it. Propagation is an observable process, not an assumption.

30. Corrections retractions and preservation events

Post-publication correction changes both the content and its status. A minor amendment, a substantial correction, an expression of concern, and a retraction communicate different judgments. They should not be reduced to a single label meaning that something changed.

Crossref launched Crossmark in 2012 to make updates and status information discoverable from scholarly content. COPE's 2019 Retraction Guidelines, Version 2, describe retraction as a means of correcting the literature and alerting readers to work that should not be relied upon. The guidance distinguishes the reliability of the record from a simple determination of authorial blame. [51, 52]

Preservation introduces a longer time scale. PREMIS version 3.0, released in 2015, models preservation information through entities including objects, events, agents, and rights. Such records can connect a digital object with actions performed on it and the responsible participants. They support the examination of preservation history rather than merely the retention of a final file. [53]

Let us consider a hypothetical article corrected after an erroneous figure was discovered. The correction notice identifies the affected figure and the consequence for the argument. A repository holds an earlier copy. A later migration changes the file format. We now have several distinct events: scholarly correction, distribution, and preservation transformation. Each needs its own relation to the object and its status.

Preserving the earlier state can be important even when it is no longer the recommended version for use. It allows investigators to determine what readers could have seen at a particular time. Deleting the evidence of the error can make the correction harder to understand. Presentation should make the status clear while retaining an appropriate historical record.

The history thus returns to a problem already visible in errata and replaced leaves. A correction does not reach every copy automatically, and a surviving copy carries a particular state. Digital systems can improve the visibility of that state if the relevant relationships are recorded and exposed. Without them, easy alteration can make the publication's history less observable.

PART VII

Diagnostic synthesis and worked investigations

31. AI assisted writing and plausible incorrectness

Generative language systems brought a new scale of assistance to drafting, rewriting, summarizing, and translation. They can also produce fluent statements whose evidential relationships are defective. The apparent completeness of a paragraph or reference list can make such problems less conspicuous than a broken sentence.

William H. Walters and Esther Isabelle Wilder's 2023 study examined 636 references generated in 84 short literature reviews on 42 multidisciplinary topics using GPT-3.5 and GPT-4. It found fabricated references as well as substantive errors in references to real works. These results concern the systems and experimental conditions studied; they should not be converted into fixed error rates for every later model or workflow. [54]

A separate 2023 study by Weixin Liang and colleagues found that tested AI-text detectors frequently misclassified writing by non-native English writers. That result is relevant to publishing diagnostics because a detector's output is an observation about a model's classification behavior, not direct evidence of authorship. [55]

Let us consider a hypothetical AI-assisted revision that changes "may indicate" to "demonstrates." The new sentence is shorter and more decisive, but it makes a stronger claim. A grammatical comparison may report an improvement while an evidential comparison identifies a regression. The review needs to track the relationship between the statement and its support.

Another example concerns a generated citation. The author name, title, journal, and date may each resemble real bibliographic material while their combination identifies no actual work. Checking only the form of the reference will miss the defect. Checking existence is necessary, and checking support for the claim remains a separate task.

The relevant diagnostic artifacts include the source passages supplied to the system, the requested transformation, the returned text, and the subsequent human changes. Retaining them where appropriate can help localize a drift. Repeating the same prompt is not guaranteed to reproduce the same text, so the actual response matters.

The historical connection is straightforward. Automated assistance changes the speed and form of error production as well as the speed of correction. Our response should be to examine the resulting claims and transformations, using evidence appropriate to each. A confident tone is a property of the prose; it is not a substitute for that evidence.

32. Comparing evidence across the publication process

We can now connect the historical cases through a common analytical question: what relationship does this artifact make available for examination? A manuscript correction exposes a change to an inscription. A proof exposes an intermediate presentation. A revision history exposes selected differences among stored states. A production log exposes recorded events. These are complementary viewpoints.

Our theoretical software diagnostics distinguishes artifacts, analysis patterns, and explanations of possible mechanisms. Applied to publishing, this encourages us to begin with the evidence actually present and to avoid moving directly from a familiar symptom to a preferred cause. The application is our comparative synthesis. It does not imply that the earlier practices originated in software diagnostics. [56]

The following distinctions help prevent a successful check at one level from being treated as proof at every level.

Examination	What it can establish	What remains separate
Character and text comparison	Recorded additions, omissions, and substitutions	Whether the change preserves meaning
Structural validation	Conformance to the encoded rules checked	Truth, rhetorical adequacy, and unencoded requirements
Bibliographic verification	Identity and metadata of a source	Whether the source supports the claim
Rendered page inspection	Visible layout under the inspected conditions	Other renderers, reading order, and physical production
Output provenance check	Relationship between a retained output and its recorded source	Approval and correctness of that source
Reader or accessibility evaluation	Observed use under stated conditions	Every possible reader and environment

Table 1. Types of examination in writing and publishing.

Consider a hypothetical statement that a publication is “verified.” Without a scope, the statement is difficult to assess. It might mean that the file opens, that references resolve, or that all claims were examined against sources. A useful verification record names the object, the examination, the result, and any unresolved condition relevant to the conclusion.

There is also a distinction between a state and a history. A final document can establish what it contains now. It may not establish how a wrong value entered it. A log can show that a conversion occurred without preserving the complete resulting page. To explain a transformation, we may need both the before-and-after states and the activity connecting them.

The synthesis therefore favors connected evidence over a single comprehensive score. We can ask which observations agree, which disagree, and which share the same source of possible error. Agreement is strongest when its dependencies are understood. Several outputs from one defective source are several manifestations of a condition, not several independent validations.

33. A repertoire of publishing analysis patterns

Our trace and log analysis pattern reference explicitly extends its scope to text, narrative, and data. Some patterns can therefore provide useful viewpoints on publishing activity. Their application requires a defined artifact and a clear correspondence. We should not turn a suggestive resemblance into an assertion that the publishing process has the same mechanism as a software execution. [57]

The table proposes applications of selected established patterns. The publishing descriptions are examples developed here, not quotations from the reference.

Analysis pattern	Possible publishing application	Question that remains
Adjoint Thread of Activity	Follow every event concerning one chapter, source, or query across participants	Is the identifier stable across systems?
Discontinuity	Examine a gap between review, approval, and production events	Was activity absent, unrecorded, or recorded elsewhere?
Message Invariant	Separate the fixed form of a repeated check report from changing values	Are equivalent messages being compared?
Counter Value	Inspect unresolved-query counts or broken-link totals over versions	Does the counter measure the intended condition?
Characteristic Message Block	Recognize a recurring group of production messages	Is the group associated with the present failure?
Significant Event	Locate an approval, source replacement, or template change	Did the event precede or merely accompany the discrepancy?
Master Trace	Compare recorded production activity with an expected sequence	Is the expected sequence valid for this publication?
Meta Trace	Examine how records and the checks that produce them change across versions	Did apparent improvement arise from changed reporting?

Table 2. Proposed applications of diagnostic analysis patterns to publishing.

An Adjoint Thread of Activity is formed by following a selected attribute or criterion across events that need not belong to one execution thread. In publishing, we might follow the identifier of a reference through author queries, copyediting, typesetting, and post-publication correction. The correspondence is useful only if records preserve that identity or provide a defensible mapping when it changes.

A hypothetical decrease in unresolved-query count illustrates the need for context. The count may fall because queries were answered. It may also fall because a file was replaced by an older copy or because a filter excluded one category. Counter Value identifies the observation; additional evidence distinguishes the mechanisms.

Meta Trace concerns the evolution of a trace or log as the software and its tracing or logging change across versions. In the publishing application proposed here, it becomes relevant when the workflow changes. A new checker may report more warnings than its predecessor while the publication itself improves. Comparing raw totals would produce a misleading conclusion. We need to compare the histories of the diagnostic instruments and their reporting rules as well as the histories of the publication.

These applications form a working repertoire, not a claim of a completed universal pattern language for publishing. Their value lies in directing attention to evidence that an ordinary linear reading can miss.

34. A worked investigation of conflicting publication outputs

The following case is hypothetical. A publisher has prepared a technical book in Word, a print file, and an EPUB. A reviewer reports three discrepancies: the EPUB reverses the meaning of one sentence, a printed cross-reference points to the wrong page, and a bibliographic link opens an unrelated article. The author's current Word file appears to contain the correct sentence and reference.

We first preserve the actual outputs reported by the reviewer and identify the Word file used for comparison. Rebuilding immediately would risk replacing evidence of the reported condition. We record the output identifiers, the stated source version, and the available production records. The question is not yet which file is right in general. It is which state and transformation account for each discrepancy.

The sentence concerns a hypothetical rule: "Do not delete the source images before checking the OCR." The reported EPUB omits "not." We compare the text in the EPUB package with the rendered display. The omission is already present in the stored content, so this instance is not explained by visual clipping. We then compare the retained intermediate text with the current Word source.

Suppose the intermediate text was exported from an earlier draft in which the omission existed. The production record shows that this intermediate file was reused after the Word correction. The mechanism is now more specific: the EPUB build consumed a stale intermediate representation. The diagnosis is supported by a chain of identifiable artifacts, not merely by the fact that the final versions disagree.

The printed cross-reference requires another investigation. In the print file, the referenced chapter begins on page 42 while the sentence says page 39. The corresponding Word field targets the correct heading, but its displayed result was not regenerated after an insertion. Here the source relationship and the displayed value have diverged. Replacing the number manually could repair this copy while leaving the maintenance mechanism unreliable.

The bibliographic link presents a third mechanism. The title and author in the reference entry identify the intended paper, but the embedded hyperlink contains the DOI of a neighboring entry. The visible text survived a copy-and-paste operation while the link target did not receive the same correction. Link resolution succeeds, yet identity verification fails. The appropriate repair must address the relationship between display text and target.

We now have three symptoms and three localized explanations. They should not be collapsed into one general claim that the export software is defective.

Reported symptom	Evidence that localizes it	Repair and follow-up
Missing negation in EPUB	The omission exists in a reused intermediate file from an earlier draft	Regenerate from the approved source and compare the affected passage and other changed passages
Wrong printed page reference	The reference targets the correct heading but retains an old display value	Update fields, regenerate the output, and check navigation in the resulting pages
Link opens the wrong article	Correct visible citation and incorrect embedded DOI target	Correct the target and verify both bibliographic identity and support for the cited claim

Table 3. Symptoms, evidence, and repairs in the worked publication case.

The repairs also have a shared verification requirement. We rebuild all outputs from the identified approved state, retain the production record, and inspect the results in their relevant representations. A text comparison checks the repaired sentence. Page inspection checks the cross-reference. Bibliographic examination checks the link and its evidential role. EPUB validation can identify structural problems but does not replace those checks.

We then examine the scope of the mechanisms. Reuse of a stale intermediate may have affected other recent corrections. An old field result may indicate other stale navigation. Copying link-bearing text may have created further mismatches. These are specific reasons to broaden the investigation. They do not justify an indefinite search for every imaginable defect.

Finally, we record which outputs were replaced and where earlier copies may remain. The correction is complete for the rebuilt artifacts, but dissemination has its own state. The case illustrates how diagnosis, repair, verification, and distribution are connected activities with distinct evidence and completion conditions.

35. Designing publications for future diagnosis

The historical cases suggest practical design choices for preserving diagnostic evidence. A publication can retain an approved source state, the decisions that materially changed it, and the relationship between that state and the distributed outputs. These records need not reproduce every keystroke. They should preserve the distinctions necessary to investigate consequential discrepancies.

Stable identification is one such choice. A chapter, figure, bibliographic item, or unresolved query should remain traceable when its position changes. A filename and a page number alone are often insufficient. Where a tool cannot preserve identity directly, a small concordance can record how old and new identifiers relate.

Another choice concerns the evidence behind claims. A reference list is valuable, but a difficult assertion benefits from a retained source location and a statement of the supported point. This makes later revision less likely to detach the source from its original role. It also helps distinguish a correction of metadata from a correction of the argument.

Production records should capture dependencies that can change the result. Templates, fonts, conversion tools, and relevant settings can affect the output even when the prose is unchanged. The appropriate amount of recording depends on the work. A simple letter and a multilingual mathematical book do not require identical diagnostic arrangements.

Checks should be selected by the failure they can reveal. A spelling pass examines selected language forms. A claim review examines support. A rendered-page review examines visible presentation. A reader evaluation examines use under stated conditions. Their results can be combined, but their scopes should remain visible.

The final choice concerns correction itself. A change should identify the discrepancy it addresses, preserve the intended meaning, and trigger re-examination of affected relationships. The scope can be local where the mechanism is local. It should expand where the evidence shows a shared dependency or a broader propagation path.

Designing for diagnosis does not mean reducing writing to compliance with a rigid process. It means making the work available for informed examination while preserving the author’s decisions and the reader’s ability to evaluate them. The same arrangements support creative revision: they make it easier to explore alternatives without losing the evidence of what changed.

Conclusion

Writing and publishing have long contained practices of comparison, correction, and reconsideration. Manuscript hands, printing-house proofs, errata, editorial reports, and digital histories preserve different evidence of those practices. Their diagnostic value depends on the relationship between the artifact and the process it is used to explain.

The modern vocabulary of observability helps us ask what a workflow makes inspectable and what it leaves unavailable. The debugging metaphor, explicitly developed for prose in Dupré's BUGS in Writing, directs attention to recurring difficulties and their correction. Software-based publishing adds literal debugging problems while retaining older questions about meaning, authority, and the identity of a text.

Our synthesis is methodological rather than a claim that one discipline historically generated all the others. We compare the operations by which practitioners establish a discrepancy, gather evidence, examine explanations, apply a change, and assess its consequences. The comparisons become useful when their limits are retained.

A publication is dependable to the extent that its claims, representations, and transformations can be examined at the levels where they may fail. We should be able to ask which text was intended, which version was used, which output was produced, and which correction reached the reader. These questions connect the history of writing and publishing with the continuing development of pattern-oriented diagnostics.

Appendix A Selected chronology

The dates identify selected works, objects, publications, and institutional developments. They do not claim the first occurrence of correction or diagnosis. A later analytical viewpoint is distinguished from the date of the evidence it examines.

Date or period	Development	Diagnostic significance and source
Late first century CE	Quintilian discusses correction in Institutio oratoria	Revision includes judgment, distance, and restraint [1]
Fourth century and later	Codex Sinaiticus and its successive corrections	One manuscript preserves multiple interventions [2, 3]
868	Dated Diamond Sutra scroll	Colophon and woodblock production distinguish text, matrix, and copy [4]
Thirteenth century	Surviving Tripitaka Koreana woodblocks	The retained production resource has its own preservation history [6]
1377	Surviving second volume of Jikji	Evidence of Korean movable-metal-type printing [5]
Fifteenth to seventeenth centuries	European printing-house correction	Correctors and authors negotiate textual production [10]
1631	Wicked Bible	Omission of a negation changes the commandment's meaning [13]
1644	Milton's Areopagitica	Publication control differs from accuracy correction [14]
1665	Philosophical Transactions begins	Editorial mediation and circulation of scholarly claims [25]
1683–1684	Moxon's account of printing	Documentary evidence of composition and correction practices [9]
1832	Systematic refereeing at the Royal Society	Reports become part of an organized evaluation process [25]
Nineteenth century	Flaubert's Madame Bovary manuscripts	Drafts preserve changes beyond local error repair [15]
1887	Lanston's Monotype invention	Mechanized composition creates new production dependencies [29]
1906	First Chicago Manual of Style	Shared conventions become a durable editorial reference [18]
1918	Strunk's The Elements of Style	Compact principles support repeated review of prose [19]
Twentieth century	Magazine fact-checking departments develop	Checking claims becomes an identifiable editorial responsibility [23]
1948	Flesch's readability paper	Selected prose features become numerical indicators [20]
1950–1951	Greg's copy-text essay	Base-text authority is distinguished from judgment of variants [7]
1961 and 1964	Selectric and magnetic-tape MT/ST	Mechanized writing is connected with retained editable state [31]
1963	Hinman's First Folio study	Optical comparison supports reconstruction of printing differences [12]

Date or period	Development	Diagnostic significance and source
Circa 1965	Lumitype 540 photocomposer represented in museum collections	Photographic composition has a distinct technical apparatus [30]
1980 and 1981	Sommers; Flower and Hayes	Revision and composing are examined as interacting activities [16, 17]
1981	Knuth and Plass on paragraph breaking	Layout decisions become explicitly algorithmic [36]
1982 and 1983	McIlroy's spelling paper; Macdonald on Writer's Workbench	Computational aids expose selected prose features [33, 34]
1984	The TeXbook	A programmable document-production system receives detailed documentation [35]
1993	Acrobat and PDF launched	Exchangeable page representation adds new production relationships [37]
1994	TEI P3 released	Structured encoding supports explicit textual relationships [38]
1995 and 1998	BUGS in Writing and its revised second edition	Debugging becomes an explicit organizing metaphor for prose [21, 22]
2007	Smith's overview of Tesseract	OCR mechanisms are documented for technical examination [44]
2008	NISO recommendations for journal article versions	Named publication states improve version identification [26]
2009	Madame Bovary manuscript website opens	Digitized manuscripts are accompanied by transcription and a genetic table [15]
2012	JATS standard and Crossmark launch	Structural interchange and publication-status discovery develop separately [39, 51]
2015	PREMIS version 3.0	Preservation objects and events are modeled explicitly [53]
2019	COPE Retraction Guidelines, Version 2	Correction of the literature receives explicit procedural guidance [52]
2023	EPUB 3.3 Recommendation; studies of generated citations and detectors	Format conformance, evidence verification, and authorship inference remain distinct [46, 54, 55]

Table 4. Selected milestones in writing and publishing.

Appendix B Glossary

Accidentals. In textual editing, features such as spelling and punctuation considered in relation to substantive readings; the distinction requires attention to the editorial theory and the actual case.

Adjoint Thread of Activity. A sequence selected by an attribute or criterion across events that need not originate in the same execution thread; applied here to a chapter, reference, query, or other publishing entity.

Apparatus. A structured presentation of readings, witnesses, and editorial information accompanying a text.

Approved source. The identified state authorized for a particular production operation; approval is separate from a claim of exhaustive correctness.

Artifact. A retained object or representation through which an investigation can proceed.

Baseline. The identified state used for comparison or further revision.

Build. A production operation that transforms sources and dependencies into one or more outputs.

Cancel. A replacement leaf or other substituted portion in a printed publication; its presence belongs to the history of the particular copy.

Collation. Comparison of textual witnesses or copies to establish differences; bibliographical usage can also concern the physical arrangement of a book.

Colophon. A statement associated with a manuscript or publication that can record information about its production, participants, or date.

Conformance. Satisfaction of specified requirements under a defined examination; it does not establish all dimensions of quality.

Copyediting. Editorial examination of text for correctness, consistency, and other matters within an agreed brief.

Copy-text. A text selected as a basis for an edition, with an authority whose extent depends on the editorial method.

Correction. An intervention intended to repair an identified discrepancy relative to a stated reference condition.

Critical edition. An edition that evaluates textual evidence and presents an editorially established text with an appropriate account of its basis.

Diagnostic artifact. An artifact retained or selected because it can support an investigation of a condition, discrepancy, or failure.

Diplomatic transcription. A transcription that preserves specified features of a particular witness rather than silently presenting a reconstructed text.

DOI. Digital Object Identifier; an identifier used within an infrastructure for persistent identification and resolution, not a guarantee of factual accuracy.

Emendation. An editorial change intended to correct or reconstruct a reading, distinguished from the direct reporting of a witness.

EPUB. A structured format for electronic publications, interpreted through reading systems and their supported capabilities.

Erratum. A published correction; practices governing the terms erratum and corrigendum vary among publishers.

Evidence taxonomy. A classification of artifacts by what they can reveal and the limits of the conclusions they support.

Fact checking. Examination of claims against relevant evidence, including identity, accuracy, context, and scope.

Genetic criticism. Study of the development of a work through drafts and other evidence of composition.

Impression. A printed result or printing run, depending on context; the intended bibliographical sense should be specified.

Invariant. A feature expected to remain unchanged under a defined set of comparisons or transformations.

ISBN. International Standard Book Number; an identifier assigned according to the rules for publications, editions, and formats.

JATS. Journal Article Tag Suite; a standardized set of elements and attributes for representing scholarly article content and metadata.

Markup. Explicit representation of selected document features and relationships for processing or interpretation.

Master Trace. An analysis pattern using an expected trace as a reference for execution in which functional and non-functional requirements are met; applied here to a specified sequence of publishing activities.

Meta Trace. An analysis pattern concerned with how a software trace or log evolves as code and its tracing or logging change across versions; applied here to changes in publishing records and the checks that produce them.

OCR. Optical character recognition; computational estimation of text from images.

Observability. Here, the arrangements through which relevant document states, relationships, and transformations become available for examination.

Preflight. Examination of a production file against a selected set of technical requirements.

PREMIS. A preservation-metadata framework that records information about digital objects and associated events, agents, and rights.

Proof. An intermediate presentation made available for examination before the relevant publication or production step.

Proofreading. Checking material prepared for publication within an agreed scope, including implementation of corrections and relevant presentation details.

Provenance. Information about origins, custody, transformations, and other relationships needed to interpret an artifact’s history.

Regression. A loss of previously satisfactory behavior or content following a change; the comparison must identify the relevant baseline and criterion.

Rendering. Production of an appearance or other presentation from stored content and processing instructions.

Retraction. A publication-status action alerting readers that work should not be relied upon in the stated respects, accompanied by an appropriate notice.

Revision. A change to a work that may repair an error, improve expression, alter structure, or change purpose.

Schema. A formal description of selected permitted structures in a document representation.

Source support. The relationship between a claim and the evidence a cited source actually provides.

Stemma. A proposed representation of relationships among textual witnesses; its adequacy depends on the transmission evidence.

Style sheet. A record of conventions, terminology, and exceptions governing a particular publication.

Substantive reading. A reading considered in relation to wording and meaning, especially in distinctions drawn by textual editors.

TEI. Text Encoding Initiative; guidelines and an associated community concerned with the representation of texts in digital form.

Version of record. A named published state within a specified publishing framework, distinguished from drafts, accepted manuscripts, and later corrected states.

Witness. A document or other source that preserves evidence of a text.

Appendix C Evidence taxonomy and blind spots

No artifact provides a complete view of a publication. The table identifies common evidence types and the assumptions that should accompany their use. It is a diagnostic classification proposed in this history.

Evidence type	Useful observations	Principal blind spot
Manuscript or draft	Visible readings, deletions, additions, and arrangement	Unrecorded intention and uncertain composition order
Editorial correspondence or query record	Reasons, requests, disagreements, and decisions	Whether the requested change was implemented correctly
Marked proof	Located corrections to a particular intermediate state	Later changes and conditions outside that proof
Printed copy	Actual text, material arrangement, and production features	Other copies and unretained production stages
Erratum or correction notice	A recognized discrepancy and proposed amendment	Unrecognized errors and copies that never receive the notice
Bibliographic record	Identifying metadata and cataloged relationships	Accuracy of every field and support for a cited claim
Source file	Editable content and stored instructions	Actual rendering and the state used for an earlier output
Revision history	Retained differences and selected event information	Unrecorded work, motives, and intermediate states
Build or conversion log	Recorded operations, warnings, and dependency information	Unlogged conditions and the complete visual result
Rendered page image	Appearance under specific rendering conditions	Searchable text, semantic structure, and other environments
OCR text	Searchable estimate of words and sequence	Misrecognition, missing text, and layout errors
Validation report	Results of the rules actually checked	Unencoded semantics, truth, and untested requirements
Reader evaluation	Observed comprehension or use in a stated setting	Unrepresented readers and circumstances
Preservation event record	Recorded transformations and participants	Unrecorded events and undocumented losses

Table 5. Publishing evidence and its principal blind spots.

C1 Questions to carry with an artifact

The questions below are intended for a concrete investigation. They help establish whether an artifact is suitable for the conclusion being drawn from it.

1. What object does the artifact represent, and which version or state does it preserve?

2. Who produced it, by what operation, and under which relevant conditions?

3. What comparison makes the observed feature a discrepancy?

4. Is that comparison testing text, meaning, presentation, provenance, or publication status?

5. Which transformations occurred between the retained source and the observed result?

6. Which parts of the proposed explanation are directly evidenced and which are reconstructed?

7. What else could produce the same observation, and what evidence would distinguish those explanations?

8. Which dependent passages, outputs, or distributed copies may be affected by the same mechanism?

9. What will demonstrate that the repair is present and that the affected relationships still hold?

10. What record will allow a later reader or investigator to understand the correction?

C2 Keeping kinds of verification distinct

Textual fidelity asks whether a representation preserves the selected source. Factual accuracy asks whether its claims agree with relevant evidence. Structural validity asks whether it conforms to specified rules. Visual review asks whether the inspected presentation is readable and correctly arranged. Accessibility evaluation asks whether information and operation remain available under relevant modes of use. Publication-status checking asks which corrections or notices apply to the particular work.

These examinations can inform one another without becoming interchangeable. A document may be structurally valid and factually wrong. A faithful transcription may preserve an error deliberately. A visually clean page may contain an inaccessible reading order. A corrected master file may coexist with uncorrected copies. An adequate diagnostic conclusion states which condition was examined and how far the evidence supports it.

References

Sources are numbered by first citation. Institutional histories, original works, research articles, standards, and technical documentation are identified separately where their roles differ. Online resources were consulted on October 3, 2026; this bibliography was updated following source reviews on October 19 and 20, 2026. URLs are printed explicitly as well as linked.

[1] Quintilian. Institutio oratoria, Book 10, chapter 4. Late first century CE. English translation by John Selby Watson (1856), edited online by Lee Honeycutt as Institutes of Oratory. Kairos archive. https://kairos.technorhetoric.net/stasis/2017/honeycutt/quintilian/10/chapter4.html

[2] Codex Sinaiticus Project. About Codex Sinaiticus. Institutional description of the manuscript and its historical significance. https://codexsinaiticus.org/en/codex/default.aspx

[3] Codex Sinaiticus Project. The transcription. Description of scribes, correctors, readings, and their digital representation. https://www.codexsinaiticus.org/en/project/transcription_detailed.aspx

[4] British Library, International Dunhuang Programme. Buddhist texts: The Diamond Sūtra. Collection account of the dated 868 scroll. https://idp.bl.uk/discover/learning/buddhism-on-the-silk-roads/articles/buddhism-on-the-ground/buddhist-texts-the-diamond-sutra/

[5] UNESCO. Baegun hwasang chorok buljo jikji simche yojeol, volume II. Memory of the World entry for Jikji, printed in 1377. https://www.unesco.org/en/memory-world/baegun-hwasang-chorok-buljo-jikji-simche-yojeol-volii-second-volume-anthology-great-buddhist-priests

[6] UNESCO. Printing woodblocks of the Tripitaka Koreana and miscellaneous Buddhist scriptures. Memory of the World entry. https://www.unesco.org/en/memory-world/printing-woodblocks-tripitaka-koreana-and-miscellaneous-buddhist-scriptures

[7] Greg, W. W. The Rationale of Copy-Text. Studies in Bibliography 3 (1950–1951), 19–36. https://www.jstor.org/stable/40381874

[8] Text Encoding Initiative. TEI P5 Guidelines, chapter 13, Critical Apparatus. Version 4.12.0, July 28, 2026 (consulted October 20, 2026). Documentation of apparatus entries, readings, and witnesses. https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html

[9] Moxon, Joseph. Mechanick Exercises on the Whole Art of Printing. Originally issued 1683–1684. Vol. 2 of the reprint edited and annotated by Theodore Low De Vinne. New York: Typothetæ of the City of New York, 1896. Project Gutenberg ebook 78991; includes the compositor and correction material. https://www.gutenberg.org/cache/epub/78991/pg78991-images.html

[10] Grafton, Anthony. The Culture of Correction in Renaissance Europe. The Panizzi Lectures 2009. London: British Library, 2011. Author’s bibliography, entry 02-015. https://graftoniana.princeton.edu/bibliography/

[11] Chartered Institute of Editing and Proofreading. What is the difference between copyediting and proofreading? Companion guidance: What is copyediting? and What is proofreading? Professional descriptions of stages, responsibilities, and the agreed brief. https://www.ciep.uk/resource/what-is-the-difference-between-copyediting-and-proofreading.html https://www.ciep.uk/resource/what-is-copyediting.html https://www.ciep.uk/resource/what-is-proofreading.html

[12] Werner, Sarah. Welcome to The Collation. Folger Shakespeare Library, August 18, 2011. Account of textual variants, optical collation, and Hinman’s comparison of First Folio copies. https://www.folger.edu/blogs/collation/welcome-to-the-collation/

[13] Folger Shakespeare Library. Manifold Greatness exhibition material. Collection descriptions, including the 1631 Wicked Bible and other corrections in printed Bibles.
https://folgerpedia.folger.edu/Manifold_Greatness_exhibition_material

[14] Milton, John. Areopagitica. 1644. Digitized text, Project Gutenberg.
https://www.gutenberg.org/ebooks/608

[15] Bibliothèques de Rouen. Manuscrits. Rotomagus. Description of the Madame Bovary manuscript collection, digitization, and the website opened in 2009.
https://www.rotomagus.fr/rotomagus/content/Manuscrits

[16] Sommers, Nancy. Revision Strategies of Student Writers and Experienced Adult Writers. College Composition and Communication 31, no. 4 (1980), 378–388. DOI: 10.58680/ccc198015930.
https://doi.org/10.58680/ccc198015930

[17] Flower, Linda, and John R. Hayes. A Cognitive Process Theory of Writing. College Composition and Communication 32, no. 4 (1981), 365–387. DOI: 10.58680/ccc198115885.
https://doi.org/10.58680/ccc198115885

[18] University of Chicago Press. About The Chicago Manual of Style. Institutional account of the manual's development and first publication in 1906.
https://www.chicagomanualofstyle.org/help-tools/about.html

[19] Strunk, William, Jr. The Elements of Style. First published 1918. Project Gutenberg ebook 37134 reproduces the Harcourt, Brace and Company text with copyright notices dated 1918, 1919, and 1920. Publication history: Lowery, George. "Omit needless words: Strunk and White's classic, The Elements of Style, turns 50." Cornell Chronicle, March 26, 2009.
https://www.gutenberg.org/cache/epub/37134/pg37134-images.html
https://news.cornell.edu/stories/2009/03/omit-needless-words-elements-style-turns-50

[20] Flesch, Rudolf. A New Readability Yardstick. Journal of Applied Psychology 32, no. 3 (1948), 221–233. DOI: 10.1037/h0057532. Article and PubMed bibliographic record.
https://doi.org/10.1037/h0057532
https://pubmed.ncbi.nlm.nih.gov/18867058/

[21] Dupré, Lyn. BUGS in Writing: A Guide to Debugging Your Prose. Addison-Wesley, 1995. ISBN 0-201-60019-6. First-edition record in the Internet Archive.
https://archive.org/details/bugsinwritinggui00dupr

[22] Dupré, Lyn. BUGS in Writing, Revised Edition: A Guide to Debugging Your Prose. 2nd edition. Addison-Wesley Professional, 1998. ISBN 978-0-201-37921-1. Publisher's record and sample content.
https://www.informit.com/store/bugs-in-writing-revised-edition-a-guide-to-debugging-9780201379211

[23] Helfand, Zach. The History of The New Yorker's Vaunted Fact-Checking Department. The New Yorker. Published online August 25, 2025; September 1 & 8, 2025 print issue.
https://www.newyorker.com/magazine/2025/09/01/the-history-of-the-new-yorkers-vaunted-fact-checking-department

[24] Crossref. Metadata Retrieval. Documentation of metadata access and reference matching.
https://www.crossref.org/documentation/retrieve-metadata/

[25] Royal Society. 360 Years of Royal Society Publishing. Institutional history identifying the journal's 1665 origin and the introduction of systematic peer review in 1832.
https://royalsociety.org/journals/publishing-activities/publishing350/publishing360/

[26] NISO and ALPSP. Journal Article Versions (JAV): Recommendations of the NISO/ALPSP JAV Technical Working Group. NISO RP-8-2008, April 1, 2008.
https://www.niso.org/publications/niso-rp-8-2008-jav

[27] Unicode Consortium. Unicode Standard Annex 15: Unicode Normalization Forms. Versioned technical specification.
https://www.unicode.org/reports/tr15/

[28] W3C. Arabic & Persian Layout Requirements. W3C Group Draft Note, October 2, 2025. Script layout requirements, including bidirectional text.
https://www.w3.org/TR/2025/DNOTE-alreq-20251002/

[29] Brignell, Hannah. Studying the Monotype Collection. Science Museum Group, December 7, 2020. Collection account including Tolbert Lanston’s 1887 invention.
https://blog.sciencemuseumgroup.org.uk/studying-the-monotype-collection/

[30] Science Museum Group. Lumitype Series 540 Photocomposer, circa 1965. Object number 1976-477, comprising parts Pt1–Pt5. Science Museum Group Collection Online.
https://collection.sciencemuseumgroup.org.uk/objects/co38012

[31] IBM. The IBM Selectric. Institutional history covering the 1961 typewriter and 1964 MT/ST.
https://www.ibm.com/history/selectric

[32] Computer History Museum. Butler Lampson. Companion collection documentation: Xerox Alto source code, “Applications,” identifying Bravo’s designers and implementers.
https://computerhistory.org/profile/butler-lampson/
https://xeroxalto.computerhistory.org/xerox_alto_file_system_archive.html

[33] McIlroy, M. Douglas. Development of a Spelling List. IEEE Transactions on Communications COM-30, no. 1 (January 1982), 91–99. Author’s publication list and paper.
https://mcilroy.cs.dartmouth.edu/pubs.html
https://mcilroy.cs.dartmouth.edu/spell.pdf

[34] Macdonald, Nina H. The UNIX Writer’s Workbench Software: Rationale and Design. Bell System Technical Journal 62, no. 6, part 3 (July–August 1983), 1891–1908. DOI: 10.1002/j.1538-7305.1983.tb03520.x.
https://doi.org/10.1002/j.1538-7305.1983.tb03520.x

[35] Knuth, Donald E. Computers and Typesetting. Author’s documentation page for TeX and METAFONT, including The TeXbook, Addison-Wesley, 1984.
https://cs.stanford.edu/~knuth/abcde.html

[36] Knuth, Donald E., and Michael F. Plass. Breaking Paragraphs into Lines. Software: Practice and Experience 11, no. 11 (1981), 1119–1184. DOI: 10.1002/spe.4380111102.
https://doi.org/10.1002/spe.4380111102

[37] Lamkin, Bryan. Evolution of the Digital Document: Celebrating Adobe Acrobat’s 25th Anniversary. Adobe, June 14, 2018. Retrospective on the June 1993 launch.
https://blog.adobe.com/en/publish/2018/06/14/evolution-digital-document-celebrating-adobe-acrobats-25th-anniversary

[38] Text Encoding Initiative. History. Institutional account including the release of TEI P3 in May 1994.
https://tei-c.org/about/history/

[39] NISO. NISO Publishes Journal Article Tag Suite JATS Standard. August 22, 2012. Announcement of ANSI/NISO Z39.96-2012.
https://www.niso.org/press-releases/2012/08/niso-publishes-journal-article-tag-suite-jats-standard

[40] Adobe. Preflight Profiles in Acrobat Pro. Companion documentation: Preflight Reports (Adobe Acrobat Pro). Profile-based checks and inventory reporting for resources including fonts and color spaces.
https://helpx.adobe.com/acrobat/using/preflight-profiles-acrobat-pro.html
https://helpx.adobe.com/acrobat/using/preflight-reports-acrobat-pro.html

[41] International Organization for Standardization. ISO 12647-7:2016. Graphic technology — Process control for the production of halftone colour separations, proof and production prints — Part 7: Proofing processes working directly from digital data. Scope record.
https://www.iso.org/standard/66426.html

[42] Microsoft. Track Changes in Word. Companion documentation: Compare document differences using the legal blackline option. Revision tracking and comparison of document versions.
https://support.microsoft.com/en-us/word/training/track-changes-in-word
https://support.microsoft.com/en-us/word/compare-document-differences-using-the-legal-blackline-option

[43] Chacon, Scott, and Ben Straub. Pro Git. 2nd edition. Apress, 2014. Chapter 2, Git Basics, section 2.2, Recording Changes to the Repository. Maintained online text.
https://git-scm.com/book/en/v2/Git-Basics-Recording-Changes-to-the-Repository

[44] Smith, Ray. An Overview of the Tesseract OCR Engine. Proceedings of the Ninth International Conference on Document Analysis and Recognition. IEEE Computer Society, 2007, 629–633. Author's research publication page.
https://research.google/pubs/an-overview-of-the-tesseract-ocr-engine/

[45] Library of Congress. Chronicling America: Technical Information. Documentation of digitized newspapers and limitations of OCR.
https://www.loc.gov/collections/chronicling-america/about-this-collection/technical-information/

[46] W3C. EPUB 3.3 Becomes a W3C Recommendation. May 25, 2023. Historical announcement of the first Recommendation publication.
https://www.w3.org/press-releases/2023/epub33-rec/

[47] W3C and DAISY Consortium. EPUBCheck. Official conformance checker documentation.
https://www.w3.org/publishing/epubcheck/

[48] W3C. EPUB Accessibility 1.1: Conformance and Discoverability Requirements for EPUB publications. W3C Recommendation, October 17, 2024.
https://www.w3.org/TR/2024/REC-epub-a11y-11-20241017/

[49] W3C Web Accessibility Initiative. Evaluating Web Accessibility Overview. Guidance on evaluation methods and limits of automated tools.
https://www.w3.org/WAI/test-evaluate/

[50] International ISBN Agency. ISBN Assignment. Guidance on publications, editions, and formats.
https://www.isbn-international.org/node/88

[51] Crossref. Crossmark Update Identification Service Launches to Alert Readers to Changes in Scholarly Content. April 27, 2012.
https://www.crossref.org/news/2012-04-27-crossmark-update-identification-service-launches-to-alert-readers-to-changes-in-scholarly-content

[52] COPE Council. Retraction Guidelines. Version 2, November 2019. DOI: 10.24318/cope.2019.1.4 (also used for later revisions). The archived Version 2 text preserved by the Council of Science Editors is the source for the 2019 discussion and chronology entry.
https://www.councilscienceeditors.org/assets/docs/retraction-guidelines.pdf

[53] PREMIS Editorial Committee. PREMIS Data Dictionary for Preservation Metadata. Version 3.0, June 2015, updated November 2015. Library of Congress.
https://www.loc.gov/standards/premis/v3/

[54] Walters, William H., and Esther Isabelle Wilder. Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT. Scientific Reports 13, no. 1 (2023), article 14045. DOI: 10.1038/s41598-023-41032-5.
https://doi.org/10.1038/s41598-023-41032-5

[55] Liang, Weixin, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou. GPT Detectors Are Biased against Non-native English Writers. Patterns 4, no. 7 (2023), article 100779. DOI: 10.1016/j.patter.2023.100779.
https://doi.org/10.1016/j.patter.2023.100779

[56] Vostokov, Dmitry. Theoretical Software Diagnostics: Collected Articles. 4th edition. OpenTask, February 2024. ISBN 978-1-912636-91-4.
https://dumpanalysis.org/theoretical-software-diagnostics-book

[57] Vostokov, Dmitry. Trace, Log, Text, Narrative, Data: An Analysis Pattern Reference for Information Mining, Diagnostics, Anomaly Detection. 5th edition. OpenTask, March 2023. ISBN 978-1-912636-58-7. Companion explanation of Meta Trace and its distinction from Master Trace: Trace Analysis Patterns, Part 82 (May 5, 2014).
https://dumpanalysis.org/trace-log-analysis-pattern-reference
https://www.dumpanalysis.org/blog/index.php/2014/05/05/trace-analysis-patterns-part-82/

