COMPUTER FORENSICS LAB

London digital forensics & e-discovery · Established 2007

Why Lawyers Should Test Search Terms Before Agreeing Them

Hit Counts, Sampling and the Two Hours That Save Six Figures

Guide 84 set out how to design keyword searches; this guide is about the discipline that makes any list safe to sign: testing it against the actual data before agreement. An untested term is a blind commitment: it may hit two hundred thousand documents at three percent relevance and consume the review budget, or hit nothing at all and quietly guarantee that responsive documents are never seen. Testing converts both risks into numbers: per-term hit counts, sampled precision, overlap analysis, null investigation and elusion checks on what the terms leave behind. It costs hours and requires no agreement from anyone. This guide explains the testing workflow for UK lawyers: what each measurement tells you, how to read the results, how to use them in the PD 57AD negotiation, and how the same discipline audits the other side's proposals before you accept them.

⏱ Estimated reading time: 24 min read

§ ABOUT THE AUTHOR

PREPARED BY COMPUTER FORENSICS LAB E-DISCOVERY TEAM

Computer Forensics Lab is a London-based digital forensics and electronic disclosure practice established in 2007. Its eDiscovery arm, **e-discovery.uk**, runs term testing as a standing stage on every search-based matter: hit analysis, precision sampling, elusion estimation and the reporting that converts test results into negotiation positions and CMC-ready evidence.

ESTABLISHED 2007 · LONDON | ISO 17025-ALIGNED PROCEDURES | SEARCH TESTING & VALIDATION | SAMPLING & ELUSION ANALYSIS | CPR PART 35 EXPERT REPORTS | FULL CHAIN-OF-CUSTODY DOCUMENTATION

§ CONTENTS

In this guide

§ 0 1 · ORIENTATION

Executive summary

THE HEADLINE POINT: NEVER AGREE A TERM YOU HAVE NOT RUN: TESTING TURNS SEARCH NEGOTIATION FROM ADJECTIVES INTO ARITHMETIC

Testing a term list is four measurements and a judgement. **Hit counts** per term, and unique hits (documents only that term finds): the volume picture, exposing both the budget-eaters and the dead weight. **Precision sampling:** a few hundred randomly selected hits per material term, reviewed quickly for responsiveness: converting "too broad" from complaint into percentage. **Null investigation:** zero-hit and near-zero terms diagnosed (wrong register, tokenisation, OCR posture per guide 84 §8) rather than accepted as proof of absence. **Elusion checking:** a random sample from the documents the whole list does *not* hit, reviewed for responsive material the terms miss: the estimate of what the strategy leaves behind, and the honest answer to "is this search reasonable". The results drive refinement (guide 84 §6's cycles) and arm the DRD negotiation: parties who exchange counts and samples settle scope in days, and the party holding numbers wins the CMC argument against the party holding opinions. All of it runs unilaterally on your own data before anything is agreed: the two hours that save six figures.

- **Counts first:** per-term and unique-hit tables expose the expensive and the useless before anyone commits.
- **Sample the material terms:** a few hundred random hits per term gives precision percentages the court can act on.
- **Diagnose every null:** zero hits is a symptom (register, syntax, OCR), never a finding of absence.
- **Test the leftovers:** elusion sampling estimates what the list misses: the completeness check reasonableness rests on.
- **Negotiate with evidence:** counts and samples exchanged shrink disputes; the protocol file records everything per guide 84 §7.

§ 0 2 · FIRST PRINCIPLES

The problem in plain English: signing a list you have never run

No lawyer would agree a settlement figure without reading it, but term lists are agreed unread against the only text that matters: the data. The Friday exchange goes "your terms are broadly acceptable"; the platform then reveals that one wildcard matches a fifth of the corpus, two terms return nothing because the client's staff never used those words, and half the list duplicates the other half's hits. Every one of those facts was knowable on Thursday, unilaterally, in hours.

The asymmetry is the point: testing is cheap, and the failure modes it prevents are not. An over-broad term becomes a six-figure review line or a humiliating renegotiation (guide 84's Example 2). An under-inclusive list becomes missed documents, a false disclosure certificate, and the application that follows. And an untested acceptance of the opponent's list imports their errors as your obligations. PD 57AD's cooperation model actually rewards the tested party twice: your own proposals arrive armoured in numbers, and your responses to theirs arrive with the sample results that make refusal reasonable. The rest of this guide is the workflow.

§ 0 3 · THE DASHBOARD

The measurements: hit counts, precision, overlap and nulls

- **Per-term hit counts:** documents (and families) each term returns against the indexed corpus: the first table, flagging outliers at both ends within minutes of the list existing.
- **Unique hits:** documents only that term finds: a term with high total hits but near-zero unique hits adds cost without adding coverage; a modest term with high unique hits is carrying weight: the pruning metric.
- **Precision (from sampling):** the fraction of a term's hits that are actually responsive: §4's method: the difference between a 190,000-hit term at 3% and a 9,000-hit term at 60% is the difference between a problem and a workhorse.
- **Aggregate and overlap views:** total documents hit by the list, pairwise overlaps, and the review estimate at current shape: the cost model that prices every proposed change.
- **The null list:** zero and near-zero terms queued for diagnosis: register mismatch (guide 84 §3), syntax and tokenisation faults, or unsearchable text (guide 86): each null resolved to a cause before the term is dropped or the absence believed.

§ 0 4 · THE SAMPLE

Sampling that stands up: method, size and reading results

- **Random, documented, reviewed fast:** samples drawn randomly by the platform (not eyeballed from the top of a sort), sizes recorded, reviewed against a simple responsive/not standard by someone who knows the issues: speed matters more than ceremony at the testing stage.
- **Size by purpose:** a few hundred documents per material term gives working precision estimates; validation-grade exercises (defending a certificate, elusion at §5) size samples for stated confidence and are documented accordingly: the point is fitness for the decision being made.
- **Family and date awareness:** samples read with family context (a hit inside an irrelevant family reads differently) and results cut by custodian and period, since a term can be precise in 2021 and noise in 2023.
- **Recording:** per-sample: draw method, size, reviewer, standard applied, result: into the guide 84 §7 protocol file, because six months later the sample is the evidence that the judgement was a judgement.
- **Reading honestly:** precision percentages carry sampling uncertainty: reported as estimates, compared as magnitudes, and never over-argued to a decimal place: the standing honesty rule, statistical edition.

§ 0 5 · THE LEFTOVERS

Elusion: testing what the terms leave behind

- **The question that matters most:** not "are the hits good" but "what responsive material sits in the un-hit set": elusion sampling draws randomly from the null set and reviews for responsiveness: the direct estimate of what the strategy misses.
- **Reading the result:** a low elusion rate across an adequate sample supports the reasonableness of the search; responsive documents found in the sample are read for *why* the terms missed them (new vocabulary, wrong channel, unsearchable text): each miss a routing instruction per guide 84 §8.
- **Iterate, then close:** vocabulary found in elusion review feeds new terms; the cycle runs until the elusion sample stops teaching: the documented stopping point that underwrites the disclosure certificate.
- **Scale by stakes:** a light elusion check suits a modest matter; a certificate likely to be attacked deserves a sized, confidence-stated exercise: proportionality applied to validation itself.
- **The concept travels:** the same leftovers-testing logic validates TAR stopping decisions (guide 89) and any culling step: learn it here, reuse it everywhere.

§ 0 6 · CASHING THE NUMBERS

Using the numbers: refinement and the PD 57AD negotiation

- **Refine on evidence:** broad-and-imprecise terms get the proximity/intersection treatment (guide 84 Example 2's rescue); zero-unique terms are pruned; null terms are fixed or replaced by their diagnosed cause: each change re-run and its delta recorded.
- **Exchange with the workings:** proposals sent with counts and sample summaries attached: the negotiation compressed because both sides argue about the same numbers instead of each other's adjectives.
- **Counter with arithmetic:** the disproportionate demand answered by its own cost model: hits × review rate × price, per term: which is what "unreasonable" looks like when a court can act on it.
- **Order-ready drafting:** agreed terms recited with their platform dialect reference and the testing appendices (guide 84 §11's wording): the list, its meaning and its evidence in one place.
- **Keep testing alive:** new custodians, new date ranges and rolling collections re-run the material terms: the numbers stay current or the protocol quietly rots.

§ 0 7 · THEIR LIST, YOUR SCRUTINY

Auditing the other side's terms

- **Run their list before accepting it:** against your own data: their terms' counts, precisions and nulls on your corpus are your acceptance decision's evidence, obtainable without asking anyone.
- **Demand their workings:** §11's requests: counts, sampling, dialect, OCR posture: a proposing party that has none has not done guide 84's method, and saying so (politely, with your own tables attached) reframes the negotiation instantly.
- **Probe the blind spots:** what covers their scans, their chat registers, their structured sources: the guide 84 §5 checklist applied to their methodology statement.
- **Watch for engineered narrowness:** terms tuned to miss (the product's internal codename absent; the key custodian's idiom unrepresented): elusion logic applied adversarially: what would their list systematically not find, and is that convenient?
- **Escalate with numbers:** challenges to their search framed as tested propositions ("your list, run against the disclosed sample, misses the following classes…") : the form of complaint that CMCs act on.

§ 0 8 · THE WIDER MAP

Source architecture: where else the evidence lives

Per this series' source-architecture discipline, adapted to method: a term list's test results are themselves evidence about where responsive material lives, and every anomaly routes somewhere. Unexplained nulls route to: the unsearchable layer (OCR posture, exception items per guide 81); the register gap (chat vocabulary, codenames: guide 63 and 84 §3's harvests); the wrong-source problem (the discussion lived in a channel or system the corpus never included: the collection-scope question reopened). Elusion finds route to: new vocabulary (fed back to terms); new custodians and channels (fed back to scope); and non-text material (routed to structured-data and analytic methods). And the testing artefacts themselves: count tables, samples, elusion results, versioned deltas: live in the protocol file as the searchable record of the search, producible when methodology is challenged and readable by the expert who inherits the matter. Testing is not a gate before the real work: it is the instrument panel telling you where the real work is.

TEST SIGNAL	TERM-LAYER RESPONSE	TEXT-REPAIR LAYER	SCOPE LAYER	STRUCTURED / ANALYTIC LAYER	PROTOCOL FILE	NET EFFECT
Unexplained null	Register harvest, syntax fix	OCR posture checked	Missing source suspected	Numbers not words	Diagnosis recorded	Null becomes cause
Six-figure hit term	Proximity / intersection	n/a	Date and custodian cuts	Ranking triage	Delta recorded	Cost becomes controlled
Elusion finds	New terms from finds	Unsearchable fraction fixed	New custodians / channels	Non-text routes	Stopping point evidenced	Misses become coverage
Zero unique hits	Prune without loss	n/a	n/a	n/a	Pruning justified	List shrinks safely
Opponent's untested list	Run it yourself	Demand posture	Probe their scope	Probe their coverage	Your audit exhibited	Their burden returned

Fallback strategy in one line: treat every test anomaly as a pointer to where the evidence actually lives: terms, text, scope or tooling: and record the routing, because the record is the defence.

§ 0 9 · IN THE WILD

Worked examples

EXAMPLE 1 · THE THURSDAY TEST THAT REWROTE THE FRIDAY LETTER

A defendant's team prepared to accept a claimant list "with minor comments" until the overnight test run came back. Of forty-one terms: one hit 218,000 documents at a sampled precision of 2.5% (a surname that was also a common word); six returned zero against a corpus known to contain the underlying events; and eleven contributed no unique hits at all. The Friday letter changed character: acceptance in principle, a counter-schedule attaching the count and sample tables, proximity refinements for the monster term, diagnosed causes for the nulls (two syntax faults, four register misses with proposed replacements), and deletion of the dead weight. The claimant's team, confronted with arithmetic rather than resistance, agreed the revised list in one call. Total testing cost: one platform evening and four hours of sampling review.

EXAMPLE 2 · THE ELUSION SAMPLE THAT REOPENED SCOPE

A claimant validated its agreed list before certifying: a sized random sample from the un-hit set. The elusion rate was low, but the finds were instructive: five responsive documents, four of which discussed the disputed pricing on a channel no term could reach: a Teams channel whose export had never been collected, referenced in emails only as "the huddle". The routing was §8's: not a term problem but a scope problem. The collection was extended (guide 66's disciplines), 3,100 channel messages entered the corpus, and the certificate: when given: recited the elusion exercise and its consequence. The opponent's later attack on search adequacy collapsed against the one paragraph showing the claimant had gone looking for its own misses and acted on what it found.

EXAMPLE 3 · THE ENGINEERED LIST, MEASURED AND RETURNED

A producing party proposed a short, plausible-looking list for its own disclosure. The receiving party ran §7's audit against the productions already exchanged: the list, applied to that sample corpus, missed 70% of the documents both sides already knew to be responsive: every miss sharing a feature: the producing party's internal codename for the project, "Larkspur", appeared in none of the proposed terms, though it appeared in most of the known documents. The audit table went to the producing party with one added term and one sentence. The reply accepted "Larkspur" without discussion. At the CMC, the episode surfaced only as a recital that terms had been "refined by agreement following testing": which is how the method usually wins: quietly, before the argument anyone could have had.

§ 10 · WHERE IT GOES WRONG

Common mistakes and technical limitations

Common mistakes

- **Agreement before arithmetic:** the Friday acceptance of the Thursday-testable list: Example 1's counterfactual.
- **Counts without sampling:** volume known, precision guessed: the 218,000-hit term's 2.5% discovered in review invoices.
- **Nulls believed:** zero hits certified as absence with the cause never diagnosed: guide 84 Example 3's trap, met here at the testing stage.
- **Elusion skipped:** the hit set polished, the un-hit set never opened: reasonableness asserted with its one direct test unrun.
- **Top-of-sort "samples":** eyeballed non-random selections defending nothing when challenged.
- **Tests undocumented:** the work done, the protocol file empty: the defence performed and then discarded.
- **Their list accepted untested:** §7 skipped: their blind spots imported as your obligations: Example 3, unread.
- **Numbers over-argued:** sampling estimates litigated to decimal places: credibility spent on false precision.

Technical limitations

- **Tests measure the indexed corpus:** exception items and unOCRed text sit outside every count: guide 81's reconciliation defines what was actually searched.
- **Precision review is judgement:** quick responsive/not calls carry reviewer variance: material decisions rest on adequate samples, not single reads.
- **Elusion estimates, it does not guarantee:** a clean sample bounds, never eliminates, missed material: stated as such in every certificate it supports.
- **Rolling data moves the ground:** counts age as collections grow: material terms re-tested per §6's last bullet.
- **Platforms count differently:** family handling and index behaviour vary: cross-platform comparisons translate before they alarm, per guide 84 §4's dialect rule.

§ 11 · INTERROGATORIES & DRAFTING AIDS

Questions to ask · Suggested wording

Ask your client

- Who can do four hours of fast sampling review this week: someone who knows the issues, available before the exchange deadline?
- Are we certifying anything soon that testing should underwrite: and does the timetable leave room for the elusion pass?
- What known-responsive documents do we hold for §7's audit corpus?

Ask your opponent

- Please provide, for each proposed term, its hit count and unique hit count against your data, together with any precision sampling undertaken and the results.
- Please state what testing, including sampling of documents not retrieved by the proposed terms, has been or will be performed to validate the proposed methodology.
- Please provide re-run counts and deltas for any revision to previously exchanged terms.

Ask your eDiscovery / forensic provider

- How fast can the full count-and-unique table run on the current corpus: and what does it flag?
- Which terms merit sampling at what sizes for the decisions we face: and who reviews?
- What would a certificate-grade elusion exercise cost here, and what confidence would it state?

SUGGESTED WORDING · TESTING PARAGRAPH FOR THE DRD / SEARCH PROTOCOL

> "Prior to agreement of search terms, each party shall run its proposed terms against its own data and exchange, per term, hit counts and unique hit counts, together with the results of precision sampling of material terms (draw method and sample sizes stated). Terms returning no or minimal results shall be diagnosed and the cause stated before removal. Following application of the agreed terms, the producing party shall undertake sampling of documents not retrieved by the terms, at a size and confidence appropriate to the matter, and shall disclose the methodology and results; vocabulary or sources identified by such sampling shall be addressed by revision of terms or scope as appropriate. All testing shall be recorded in the search protocol file, and revisions to agreed terms shall be made only as exchanged, versioned amendments with re-run results."

§ 12 · QUICK CONTROL

Checklist and red flags · When to involve a digital forensic expert

The testing checklist

- ☐ Full count and unique-hit table run before any exchange
- ☐ Material terms precision-sampled: random draws, sizes recorded
- ☐ Every null diagnosed to a cause: register, syntax, text, scope
- ☐ Cost model built: hits × rate × price per term
- ☐ Refinements re-run with versioned deltas
- ☐ Elusion sample drawn, reviewed, and its finds routed
- ☐ Stopping point documented for the certificate
- ☐ Opponent's list run against own data before acceptance
- ☐ Their workings demanded; their blind spots probed
- ☐ Everything in the protocol file, dated and versioned

Red flags

- Exchange deadlines pressuring acceptance of untested terms: Example 1's setup.
- Proposals arriving without counts: the method's absence, visible.
- Certificates over undiagnosed nulls.
- Opposing lists that miss known-responsive documents: Example 3's measurement.
- "Samples" that were sorted browses.
- Elusion never mentioned in a methodology statement.
- Term revisions appearing without deltas.

When to involve a digital forensic expert

At the testing pass itself, where counts, samples and diagnostics run overnight against the live corpus (Example 1's evening); at validation, where elusion exercises are sized, run and reported to certificate grade (Example 2's paragraph); at the audit, where opposing lists are measured against known-responsive corpora (Example 3's table); and at challenge, where search adequacy is attacked or defended with the protocol file as the exhibit, under CPR Part 35 where required. Through **e-discovery.uk**, testing is a standing stage of every search matter: the instrument panel installed before the aircraft is argued about.

§ 13 · COMMON QUESTIONS

Frequently asked questions

How long does proper term testing actually take?

The count tables run in hours on an indexed corpus; material-term sampling is typically a few focused reviewer-days at most; a working elusion pass adds another. Against review budgets and renegotiation costs it rounds to free: Example 1's whole exercise fitted between a Thursday and a Friday letter.

Can we test before the DRD is agreed: is that allowed?

Not only allowed but expected: testing runs unilaterally on your own data and needs nobody's consent, and PD 57AD's cooperation duty is served, not offended, by arriving at the negotiation with evidence. What you exchange is a choice; that you tested should not be.

What is a "good" precision figure for a search term?

There is no universal number: a 60% term is a workhorse, a 2% term is a problem unless its hits are few or its unique finds are vital: the judgement weighs precision against volume, unique contribution and the issue's importance, which is why the table shows all three. The discipline is having the figures; the art is reading them together.

Do we have to disclose our test results to the other side?

Strategy varies: counts and sample summaries are routinely exchanged because they shorten negotiations and build credibility (Example 1's letter), while full protocol files usually surface only if methodology is challenged. What is unwise is having nothing to disclose: the untested party's position, under PD 57AD's cooperation lens, reads as the unreasonable one.

Our elusion sample found responsive documents. Is the search indefensible?

The opposite, if you act: elusion exists to find misses while they are fixable: the finds are routed (new terms, new sources, per §8), the cycle re-runs, and the eventual certificate recites the exercise: Example 2's sequence. Indefensibility is the mirror image: the party that never looked, explaining later why it preferred not to know.

The other side refuses to test and insists on their list. Now what?

Run their list yourself (§7), against your data and the known-responsive corpus, and put the results in correspondence: their untested insistence against your measured counter is exactly the record a CMC decides on, and usually the insistence softens before it gets there: Example 3's one-sentence ending. The refusal to engage with arithmetic is itself evidence, and courts read it fluently.

§ 14 · REFERENCE

Glossary

Confidence level
The stated reliability of a sized sample's estimate.

Cost model
Hits × review rate × price: the per-term budget arithmetic.

Elusion
Responsive material in the un-hit set; sampled to estimate misses.

Hit count
Documents a term returns against the indexed corpus.

Null set
Documents no term retrieves; the elusion sample's population.

Precision
The responsive fraction of a term's hits.

Random draw
Platform-selected sampling; the defensible kind.

Stopping point
The documented decision that testing cycles are complete.

Unique hits
Documents only one term finds; the pruning metric.

Versioned delta
A recorded, re-run change to an exchanged list.

Sources and authoritative references

The testing disciplines in this guide draw on materials published by our laboratory and e-discovery practice, referenced here as sources:

- e-discovery.uk, EDRM Phase 04 · Processing and Phase 05 · Review (term testing, sampling and validation as practised): e-discovery.uk/edrm/processing · e-discovery.uk/edrm/review
- cflab.uk, digital forensics services (search validation, methodology auditing, CPR Part 35 evidence): cflab.uk/digital-forensics-services · guides library: cflab.uk/guides
- e-discovery.uk, practice method and Knowledge Centre: e-discovery.uk · e-discovery.uk/knowledge
- PD 57AD and the Disclosure Review Document: justice.gov.uk, PD 57AD · CPR Part 35: justice.gov.uk, Part 35
- ICO, UK GDPR guidance: ico.org.uk · ISO/IEC 27037: iso.org, 27037

§ HOW A SPECIALIST LABORATORY CAN ASSIST

Working with Computer Forensics Lab

Term testing at **e-discovery.uk** is the standing stage between drafting and agreement: overnight count tables, sampled precision on the material terms, null diagnosis, the cost model, and the elusion pass sized to the certificate the matter will one day give: Examples 1 and 2 as ordinary workflow, Example 3's audit available whenever an opposing list arrives looking too tidy. Results land as negotiation-ready schedules and a maintained protocol file, and where methodology becomes the battleground the laboratory evidences it under CPR Part 35. Conflicts checks, confidential scoping discussions and fixed-fee estimates are routine; and if a term list is sitting in your inbox awaiting Friday's agreement, the test run is tonight's job: the two hours, before the six figures.

INSTRUCT THE LAB

Speak to a forensic examiner, not a salesperson.

Describe your matter in confidence and an examiner will respond the same working day. Conflicts checks and scoping calls with US counsel are routine.

NEW ENQUIRIES
+44 (0)20 7164 6971

EMAIL
info@cflab.uk

E-DISCOVERY
e-discovery.uk

Computer Forensics Lab Ltd · Euro House, 133 Ballards Lane, Finchley, London N3 1LJ, United Kingdom
cflab.uk · Customer service +44 (0)20 7164 6915 · ICO ZB395023 · NPCC · ISO 17025 aligned practice · UK Gov Digital Marketplace