COMPUTER FORENSICS LAB

London digital forensics & e-discovery · Established 2007

The Electronic Discovery Reference Model for UK Litigation Teams

Identification to Production, Through a Worked Commercial Dispute

Identification, preservation, collection, processing, review, analysis and production, each stage explained through a realistic England & Wales commercial dispute, with the volumes, deliverables and defensibility checkpoints that answer the question every litigator eventually asks: what actually happens to the evidence between collection and production?

⏱ Estimated reading time: 30 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**, delivers every phase of the EDRM in-house, forensic collection, processing and culling, hosted review and court- or regulator-ready production, with each phase closing on its own defensibility checkpoint: what was done, by whom, when, against which hash, under whose instruction. Our examiners also produce CPR Part 35 and CrimPR Part 19 compliant expert reports and give evidence in the Crown Court, County Court, High Court and employment tribunals.

ESTABLISHED 2007 · LONDON | ISO 17025-ALIGNED PROCEDURES | ISO 27001-ALIGNED SECURITY | ACPO / NPCC DIGITAL EVIDENCE PRINCIPLES | RELATIVITY-READY HOSTED REVIEW | CPR PART 35 EXPERT REPORTS | FULL CHAIN-OF-CUSTODY DOCUMENTATION

§ CONTENTS

In this guide

§ 0 1 · ORIENTATION

Executive summary

The **Electronic Discovery Reference Model (EDRM)** is the industry-standard map of how electronically stored information travels from a client's systems to a disclosure production: **identification → preservation → collection → processing → review → analysis → production**, book-ended by information governance before the dispute and presentation at trial. It is a reference model, not a rulebook, but it is the vocabulary in which providers quote, the DRD is negotiated, and disclosure methodology is defended, so litigation teams need to speak it fluently.

This guide walks the model stage by stage through a realistic Business and Property Courts dispute, *Halworth v Brydon*, a £9m supply-agreement claim, with the actual volumes, decisions and deliverables at each stage, so the abstractions become concrete: what a data map looks like, what processing does to 2.1 million items, why only 31,000 of them were ever read by a lawyer, and how 9,650 documents were produced without anything being "lost".

- **The stages are checkpoints, not ceremony.** Each phase should close with its own deliverable, data map, hold register, hash manifest, processing and exception reports, review QC statistics, production manifest. Those documents are the audit trail behind the disclosure certificate.
- **The funnel is the economics.** Costs concentrate in review; savings are made in processing. Every defensible reduction upstream of review is worth multiples of its cost downstream.
- **Nothing is deleted between collection and production.** Culling suppresses; it does not discard. Every document excluded from review or production remains in the processed population, with a logged reason, reversible on demand, which is precisely why a well-run funnel survives challenge.
- **The model is iterative in practice.** Review feeds back into search terms; analysis reopens identification (the custodian nobody mentioned); late sources re-enter at collection. The stages describe the journey of any given document, not a one-way calendar.

§ 0 2 · THE MAP

What the EDRM is, and how it maps onto English disclosure

The EDRM originated in the United States in 2005 as a community reference model and has become the global lingua franca of electronic discovery. It describes nine phases; this guide concentrates on the seven that litigation teams manage directly, with the two book-ends noted for completeness:

Figure 1 · The EDRM. Information governance (what the client did before any dispute) and presentation (trial) frame the seven phases this guide works through. Preservation and collection run in parallel in practice; review and analysis inform each other continuously.

Because the model is American in origin, two translation notes matter for England & Wales. First, vocabulary: what the model calls "discovery" is, procedurally, **disclosure**, under PD 57AD in most Business and Property Courts claims, or CPR Part 31 with PD 31B elsewhere, and the model's phases slot directly into those regimes: identification and preservation discharge the express preservation duties; collection, processing and review are the "reasonable and proportionate search"; production is disclosure and inspection in the agreed format. The Disclosure Review Document's Section 2 is, in effect, a form asking the parties to describe their EDRM. Second, culture: the English overlay adds duties the US model assumes less directly, known adverse documents under PD 57AD, the disclosure certificate's personal weight, and the court's active management of scope through Issues for Disclosure and Models A–E.

One more framing point before the worked example: the phases are managed by different people. Identification and preservation are lawyer-led with client IT; collection is specialist-led under lawyer instruction; processing is provider-led against a written protocol; review is lawyer-owned; analysis is shared; production is provider-executed and lawyer-certified. Knowing who owns which checkpoint prevents the commonest failure, everyone assuming the audit trail is someone else's job.

§ 0 3 · THE CASE

The worked example: Halworth v Brydon

The dispute that runs through this guide is fictional but assembled from the ordinary anatomy of B&PC supply disputes. **Brydon Marine Systems Ltd** (claimant) sues **Halworth Components Ltd** (defendant, our client) for £9.2m: Brydon alleges that Halworth supplied actuator assemblies that departed from the agreed specification, causing warranty claims and a lost fleet contract; Halworth's defence is that Brydon's own engineers requested the specification changes in 2023, informally, and partly over WhatsApp, and that the losses flow from Brydon's late redesign, not the components. Proceedings are issued in the Technology and Construction Court; PD 57AD applies; the parties are heading for Extended Disclosure with Model D sought on the specification and causation issues.

Halworth's data landscape, established in the first week, is typical of a 300-employee manufacturer:

SYSTEM / SOURCE	WHAT IT HOLDS	FIRST-WEEK STATUS
Microsoft 365 tenancy	Exchange mail (all staff), Teams chat and channels, SharePoint / OneDrive project sites	90-day deletion policy on Teams chat, suspended day 1 by litigation hold
Engineering file server	CAD drawings, specification revisions, test data back to 2016	Stable; snapshot schedule documented
ERP (production/quality)	Works orders, QC results, non-conformance reports	Structured data, export by report, not by "documents"
WhatsApp, commercial director	Thread with Brydon's chief engineer, 2022–2024, incl. the disputed change requests	Disappearing messages off; handset forensically extracted day 3
Departed engineer (left 2024)	Mailbox archived; laptop in IT storage awaiting reissue	Reissue stopped day 1; laptop forensically imaged day 4
Legacy Lotus Notes archive	Pre-2018 correspondence	Flagged as disproportionate; parked pending DRD negotiation

Nine custodians are initially proposed; six survive scoping (the commercial director, sales manager, two engineers including the leaver, the quality manager, the managing director). Date range: 1 January 2022 – 30 June 2025. This is the landscape the seven stages now work through.

§ 0 4 · STAGE 1

Identification

What happens: the universe of potentially relevant ESI is mapped before anything is copied: which systems, which custodians, which date ranges, what volumes, who controls each source (including third parties), and what is fragile. The method is interviews, client IT for the systems view, the custodians themselves for the shadow view (the WhatsApp thread, the personal OneDrive, the spreadsheet "everyone actually uses"), cross-checked against retention policies and any data map information governance has left behind.

The lawyer's role: lead the interviews with a checklist, ask expressly about messaging apps, personal devices and leavers, and convert the answers into two documents: the **data map** (source → custodian → location → volume → volatility → decision) and the first draft of DRD Section 2. Under PD 57AD this stage is also where Initial Disclosure's key documents are located.

Checkpoint deliverable: the data map, dated and versioned, the exhibit behind every later "why wasn't X searched?" question.

> HALWORTH V BRYDON · IDENTIFICATION
>
> The IT manager's system list missed both sources that later decided the case: the commercial director's WhatsApp thread (surfaced by his custodian interview, question 7: "do you ever deal with Brydon off email?") and the departed engineer's laptop sitting in a storeroom. The data map recorded 14 sources; 11 went forward, the Lotus Notes archive was parked with reasons, and two personal devices were noted as BYOD requiring consent. Estimated in-scope volume: ~460 GB.

Common failure: interviewing IT but not custodians, the systems view never contains the shadow IT where informal (and therefore candid) communication lives.

§ 0 5 · STAGE 2

Preservation

What happens: everything on the data map is frozen against alteration and deletion: written legal-hold notices with acknowledgment; suspension of auto-deletion (mail retention, Teams chat expiry, recycle-bin emptying); litigation holds applied in-platform (Microsoft Purview / Google Vault); IT operations told to stop reissuing, reimaging and disposing; third parties holding client data notified. PD 57AD para 4 makes each element an express duty, and requires the legal representatives to obtain written confirmation that it has been done.

The lawyer's role: issue the hold on day one, before scoping is finished, because the duty does not wait, then maintain the **hold register**: recipients, acknowledgments, systems suspended, reminder dates, releases at matter end.

Checkpoint deliverable: the hold register plus the client's written preservation confirmation.

> HALWORTH V BRYDON · PRESERVATION
>
> The hold went to 31 people (six custodians plus IT, records and the board) on day 1; the Teams 90-day chat expiry was suspended the same afternoon, recovering, as it turned out, four months of chat that would otherwise have aged out before the CMC. The storeroom laptop was pulled from the reissue queue with two days to spare. One near-miss made the file note: marketing's routine mailbox archive job ran on day 2 before its exclusion was processed, and the hold register records the incident, the assessment (no in-scope custodians affected) and the fix.

Common failure: a hold that reaches custodians but not IT *operations*, deletion is mostly done by systems and administrators, not by the people who received the notice.

§ 0 6 · STAGE 3

Collection

What happens: preserved sources are copied into the disclosure workflow by methods matched to each source's risk profile: platform exports under written protocol for routine corporate sources; targeted or full forensic acquisition for mobiles, leavers' devices and anything volatile or suspect; report-based exports for structured ERP data. Every collection is hash-verified at capture and logged, source, method, tool, version, operator, date, so the chain of custody starts here and never breaks.

The lawyer's role: decide the method per source (consciously, recorded, per the triage principles in our companion guide *eDisclosure or Digital Forensics?*), instruct the provider in writing, and prohibit custodian self-collection.

Checkpoint deliverable: the collection log and hash manifest, the documents that answer "how do we know this is complete and unaltered?"

> HALWORTH V BRYDON · COLLECTION
>
> Purview exports collected the six mailboxes, Teams data and the two project SharePoint sites (388 GB, logs retained). The engineering server was collected by targeted forensic copy of the actuator project trees (41 GB, hash-manifested). The commercial director's handset received a file-system extraction, WhatsApp database, media and metadata intact, and the leaver's laptop a full forensic image, both under laboratory chain of custody, because each carried a plausible forensic question. ERP data came out as four agreed reports rather than a database dump. Total collected: 447 GB · 2,140,000 items.

Common failure: "helpful" collection, custodians forwarding what seems relevant, IT drag-and-dropping folders, which destroys metadata, defeats verification and hands the opponent a methodology attack.

§ 0 7 · STAGE 4

Processing: what actually happens to evidence between collection and production

Processing is the least visible stage to lawyers and the one clients most often ask about, usually in the form of this guide's headline FAQ: *what actually happens to the evidence between collection and production?* The answer, in order:

- **Ingest and indexing.** Containers (PSTs, zips, forensic images' live files) are expanded, embedded objects extracted, text and metadata pulled into a database, everything full-text indexed. Items the tools cannot open, encrypted, corrupt, exotic formats, go to an **exception report** to be resolved or accepted, never silently dropped.
- **De-NISTing and system-file removal.** Known operating-system and application files are removed by hash reference, noise, not evidence.
- **De-duplication.** Exact duplicates are identified by hash (globally or per custodian) and suppressed, *family-aware*, so a duplicate parent never orphans its attachments.
- **Email threading and near-duplicate grouping.** Conversations collapse to their inclusive messages; near-identical drafts group so reviewers see one canonical version. The underlying messages are suppressed from review, not deleted.
- **OCR and language identification.** Image-only documents become searchable, with quality thresholds recorded; non-English material routes to the right reviewers.
- **Culling: dates, custodians, search terms, analytics.** The agreed date range and any negotiated search terms are applied with **hit reports at each step**, terms iterated against samples rather than run once; concept clustering and, at volume, technology-assisted review prioritise what remains.

THE PRINCIPLE THAT MAKES THE FUNNEL DEFENSIBLE

Every reduction is a **suppression, not a deletion**. The full collected population remains intact and hash-verified; each excluded item carries a logged reason (system file, duplicate of X, outside date range, no term hit); every suppression is reversible; and the volumes in and out of each step are reported. When the opponent asks what happened to the other two million items, the answer is a set of documents, produced on request, which is why a properly run funnel is an asset at the CMC rather than a vulnerability.

The lawyer's role: set the processing protocol in writing (dedupe scope, date filters, term lists, exception handling), read the hit and exception reports, actually read them, and test search terms empirically before agreeing them with the other side.

Checkpoint deliverables: processing report (volumes in/out per step), exception report with resolutions, hit reports per term, OCR quality report.

HALWORTH V BRYDON · PROCESSING

2,140,000 collected items entered processing. De-NISTing removed 512,000 system files from the laptop image's live set; family-aware deduplication suppressed a further 655,000 exact duplicates; date filtering (2022–mid-2025) suppressed 421,000; the negotiated 38-term search list, iterated three times after sample testing showed "spec" alone hit 190,000 items, left **167,000 documents plus families** as the review population. The exception report listed 1,240 items: 1,180 resolved (passwords obtained, formats converted), 60 accepted as irrecoverable with reasons logged. Every figure above went into the DRD and none was disputed.

Common failure: processing run as a black box, default settings, unread hit reports, exceptions logged to a file nobody opens, discovered only when review overruns or the opponent samples the suppressed population.

§ 0 8 · STAGE 5

Review

What happens: lawyers (and machines) read the processed population and decide, document by document: relevant to which Issue for Disclosure; privileged in whole or part; confidential; needing redaction. Review runs in a hosted platform with role-based access and full audit logging, under a written **review protocol** that defines the issues, coding rules, privilege guidance and escalation route. First-level review is prioritised by TAR/CAL where volume justifies it; second-level and privilege QC sample and correct; statistical sampling of the unreviewed remainder validates stopping.

The lawyer's role: own the protocol, calibrate early (the first 500 documents teach the coding rules), monitor QC statistics weekly, and keep the privilege log contemporaneous rather than reconstructing it at production.

Checkpoint deliverables: review protocol, QC and sampling statistics, privilege log, decision log for contested calls.

> HALWORTH V BRYDON · REVIEW
>
> Continuous active learning was trained on a 2,000-document seed; reviewers worked the ranked stream and coded 31,000 documents before the richness curve flattened; a 1,500-document random sample of the unreviewed remainder found a residual relevance rate of 0.4%, documented and accepted as the stopping justification. Output: 9,650 disclosable documents, 610 privileged (logged), 74 requiring redaction (third-party pricing and personal data). The WhatsApp thread, reviewed natively with its metadata, produced the eleven messages on which the defence now rests.

Common failure: review begun before the protocol is settled, so 20,000 documents are coded to rules that then change, the most expensive avoidable cost in the model.

§ 0 9 · STAGE 6

Analysis

What happens: analysis is the intelligence layer that runs alongside review rather than after it: communication mapping (who talked to whom, when the pattern changed), timeline construction from metadata, concept clustering to surface themes the term list missed, gap detection (the fortnight with no mail, the thread with a missing reply), and targeted deep-dives on key documents. It is where the case theory meets the data, and where forensic questions surface: the anomalous PDF, the suspiciously clean custodian.

The lawyer's role: direct analysis with questions, not curiosity ("show me every route the 2023 change request could have travelled"; "does anything corroborate the WhatsApp instruction?"), feed results back into search terms and custodian scope, and route anomalies down the forensic escalation path rather than past it.

Checkpoint deliverables: chronologies and communication maps tied to document IDs; a log of scope changes analysis triggered; forensic escalation notes where raised.

> HALWORTH V BRYDON · ANALYSIS
>
> Communication mapping showed Brydon's chief engineer moving the specification conversation off email and onto WhatsApp in March 2023, three weeks before the first disputed change, and a timeline built from file-server metadata matched each WhatsApp instruction to a CAD revision within 48 hours. Gap detection flagged that Brydon's disclosure contained no internal mail for the fortnight around its own redesign decision; a targeted Model C request followed, and the late-produced documents corroborated Halworth's causation case. One anomaly, a Brydon test report whose creation date post-dated its signature, was escalated and is now the subject of a CPR 32.19 notice.

Common failure: treating analysis as a luxury for big cases. Communication maps and metadata chronologies are cheap; trials lost to an unnoticed pattern are not.

§ 10 · STAGE 7

Production

What happens: the disclosable set leaves the building in the agreed form: unique IDs assigned; families intact; redactions burned in (with underlying text removed, not covered); natives for spreadsheets, chat and media; images-plus-text for correspondence; the agreed metadata fields in a load file (DAT/OPT or platform exchange formats); the whole production hash-manifested and quality-checked before release. Inspection follows in the agreed manner, and the production set is locked immutable so every later question can be answered against it.

The lawyer's role: agree the production specification early (it belongs in the DRD, not in correspondence the week before exchange), QC a sample of every production, redactions especially, and sign the disclosure certificate against the checkpoint documents the previous six stages have generated.

Checkpoint deliverables: production specification, production manifest with hashes, redaction log, the signed certificate.

> HALWORTH V BRYDON · PRODUCTION
>
> 9,650 documents were produced: correspondence as searchable PDF with extracted text and 14 agreed metadata fields; 512 spreadsheets and the ERP reports in native; the WhatsApp thread as both a native database export and a paginated, metadata-annotated rendering the tribunal could actually read. Pre-release QC sampled 400 documents and caught two redaction misses, logged, fixed, resampled. The manifest's SHA-256 hashes were exchanged with the production, and when Brydon later queried a document's completeness, the answer took one email: family reference, hash, collection log entry.

Common failure: format agreed late and by attrition, producing spreadsheets as PDFs (destroying formulae), breaking families, or "redacting" with rectangles over live text.

§ 11 · THE NUMBERS

The funnel in numbers: Halworth v Brydon end to end

The whole model, one table, this is the exhibit to keep for the client who asks where the money went, and for the opponent who asks where the documents went:

STAGE	ITEMS IN	ITEMS OUT (ACTIVE)	WHAT CHANGED, AND WHERE THE EXCLUDED ITEMS ARE
Collection	n/a	2,140,000	447 GB captured across 11 sources, hash-manifested; nothing yet excluded
De-NIST	2,140,000	1,628,000	512,000 known system files suppressed by hash reference, logged, reversible
De-duplication	1,628,000	973,000	655,000 exact duplicates suppressed family-aware; each mapped to its retained original
Date filter	973,000	552,000	421,000 items outside 2022 – mid-2025 suppressed; rule recorded in the protocol and DRD
Search terms	552,000	167,000	38 negotiated terms, three tested iterations, hit reports exchanged; non-hits suppressed, sampled for validation
Review (CAL)	167,000	31,000 read	Ranked review to a flattened richness curve; 0.4% residual rate in a 1,500-item sample of the unread remainder, documented
Production	31,000 read	9,650 produced	Relevant, non-privileged set; 610 privileged logged, 74 redacted; manifest hashed and exchanged

Figure 2 · From 2.14 million items to 9,650 produced documents. At every step the "missing" items are suppressed with logged reasons in an intact, verifiable population, the difference between a funnel and a shredder.

Two readings of the table repay attention. Economically: the steps that removed 92% of the volume cost a small fraction of what reviewing it would have; processing is where budgets are made. Forensically: the same table is the completeness defence, each transition is documented, reversible and sampled, which is why none of the reductions could be characterised as suppression of evidence.

§ 12 · CROSS-CUTTING DUTIES

Integrity, privilege, data protection and proportionality across the model

- **Integrity is a thread, not a stage.** Hashes taken at collection are re-verified at ingest, carried per document through processing and review, and reproduced in the production manifest, one unbroken line from the custodian's mailbox to the trial bundle. Any stage that cannot show its link in that chain is the stage the cross-examination will find.
- **Privilege runs from review to production, and backwards.** Screens (law-firm domains, named lawyers) triage; human QC at family level decides; the privilege log is built as coding happens; production QC catches the privileged attachment on the innocent covering email. Agree clawback terms in the DRD so inadvertent disclosure (CPR 31.20) has a contractual mechanism as well as a procedural one.
- **Data protection shadows every stage.** Identification and collection define what personal data is taken (minimisation: targeted trees, not whole servers); processing and hosting define where it sits and who reaches it (UK residency, encryption, role-based access); production defines what leaves (redaction of third-party personal data where irrelevant); disposal, the model's quiet final duty, returns or verifiably destroys the estate when the matter ends.
- **Proportionality is enforced through the checkpoints.** Custodian counts, date ranges, parked legacy sources, term negotiations, TAR adoption, staged disclosure, every proportionality lever in PD 57AD is pulled at a specific stage of this model, and the stage's deliverable is the evidence that it was pulled reasonably.

§ 13 · WHERE IT GOES WRONG

Common mistakes and technical limitations

Common mistakes across the model

- **Stages run out of order:** collection before the hold has landed; review before the protocol exists; production format negotiated after review has coded to a different assumption.
- **Checkpoints skipped:** no data map, no hit reports read, no sampling of the unreviewed remainder, each a gap that must later be explained from memory.
- **The shadow sources missed:** chat apps, personal devices, the leaver's laptop, identification's classic omissions, discovered by the opponent rather than the client.
- **Structured data treated as documents:** exporting an ERP as two million rows instead of four agreed reports, unreviewable, and unnecessary.
- **The funnel run as a shredder:** reductions without logs, suppressions without reversibility, turning defensible culling into an accusation.
- **The iteration ignored:** analysis findings never fed back into terms and custodians, so the exercise answers the questions asked in month one, not the case as it stands at trial.

Technical limitations to understand

- **Search recall is never 100%:** OCR errors, misspellings, code words and images of text evade terms, which is why sampling, not certainty, is the validation standard.
- **Export tools have documented gaps:** platform exports omit certain metadata, versions and recoverable items; the collection log should record what the chosen tool does not capture.
- **TAR ranks; it does not warrant:** its defensibility is the statistics around it, not the algorithm inside it.
- **Chat and mobile data resist pagination:** threads, reactions, edits and disappearing messages need native-plus-rendered production and, where contested, forensic extraction, a screenshot is neither.
- **Structured and cloud systems answer questions, not searches:** what you can get from an ERP or a SaaS platform is constrained by its reports, APIs and log-retention windows, scope with the system's owner, early.

§ 14 · INTERROGATORIES

Questions to ask

Ask your client

- Who can walk us through the systems estate, and separately, which custodians will tell us how work actually gets done?
- What auto-deletion is running right now on mail, chat and collaboration platforms, and who can suspend it today?
- Which relevant people have left, and where are their accounts and devices at this moment?
- What structured systems (ERP, CRM, finance, quality) touch the issues, and what reports can they produce?
- Is there business WhatsApp / Signal / personal-device use we need to preserve, and are those custodians ready for that conversation?
- Who will be our single point of contact for IT actions, and who confirms in writing that preservation steps are done?

Ask your opponent (through the DRD and correspondence)

- Describe your identification exercise: sources considered, included, excluded, with reasons for the exclusions.
- When did your hold go out, to whom, and what automated deletion was suspended, when?
- Collection methods per source; logs and hash manifests available?
- Your processing figures: volumes in and out at each step, hit reports per term, and the exception report.
- Review methodology: TAR or linear, validation sampling, and the statistics behind your stopping decision.
- Production: format, metadata fields, family handling, redaction basis logging, agreed now, in the DRD.

Ask an eDiscovery provider

- Which EDRM phases are delivered in-house, and where do hand-offs occur, and how does the chain of custody survive them?
- What are your per-phase deliverables, data map support, collection logs, processing and exception reports, QC statistics, production manifests, and can we see specimens?
- How are suppressions logged and reversed? Show us what "nothing is discarded" means in your workflow.
- Platform, hosting jurisdiction, security accreditations, access controls, audit logging?
- Pricing per phase, collection per source, processing per GB, hosting per GB/month, review support, and the assumptions in the estimate?
- Who from your team would give factual evidence about methodology if it were challenged, and has anyone had to?

§ 15 · DRAFTING AIDS

Suggested wording for instructions

PROCESSING PROTOCOL (CORE OPERATIVE PARAGRAPHS)

> "Processing is to proceed as follows: (1) ingest all collected data with full extraction of text, metadata and embedded objects, logging all exceptions for resolution and reporting; (2) remove known system files by NIST hash reference; (3) de-duplicate by MD5/SHA-256 [globally / per custodian], family-aware; (4) thread email and group near-duplicates, suppressing (not deleting) non-inclusive items; (5) OCR image-only documents, recording quality thresholds; (6) apply the date range [ ] and the search terms at Schedule 1 only after sample testing, providing hit reports per term per iteration. All suppressions are to be logged with reasons and reversible; volumes in and out of each step are to be reported. No collected data is to be deleted."

PRODUCTION SPECIFICATION (FOR THE DRD)

> "Documents are to be produced with a load file containing per document: unique ID; family references; custodian; source; date sent/created; author/sender; recipients; subject/filename; file type; MD5 or SHA-256 hash. Emails and standard documents: searchable PDF [or single-page TIFF] with extracted text. Spreadsheets, databases, chat exports, audio and video: native format. Chat data additionally as a paginated rendering with message-level metadata. Redactions burned in with underlying text removed and the basis logged. Families produced together. A hash manifest of the production is to be exchanged with each tranche. Inadvertently produced privileged material is subject to the clawback terms at [ ]."

WEEKLY CHECKPOINT REPORTING (PROVIDER INSTRUCTION)

> "Throughout the exercise you will provide a weekly written checkpoint report recording: work performed by phase; volumes in and out of each processing step; exceptions raised and resolved; review throughput and QC statistics; any deviation from the agreed protocols, with reasons; and any matter which in your professional judgment raises an integrity, completeness or forensic concern, which is to be reported within one working day of identification."

§ 16 · QUICK CONTROL

Checklist and red flags

The EDRM checkpoint checklist

- ☐ Identification: data map completed from IT *and* custodian interviews; shadow sources (chat, BYOD, leavers) expressly addressed; exclusions reasoned
- ☐ Preservation: hold issued day one with acknowledgments; auto-deletion suspended in-platform; IT operations on notice; hold register live; client written confirmation obtained
- ☐ Collection: method chosen per source and recorded; hash manifest and collection logs received; no self-collection
- ☐ Processing: written protocol; hit, exception and OCR reports received and read; suppressions logged and reversible
- ☐ Review: protocol settled before coding starts; TAR validated by sampling; privilege log contemporaneous; QC statistics monitored
- ☐ Analysis: chronology and communication mapping run; findings fed back into scope; anomalies escalated down the forensic path
- ☐ Production: specification agreed in the DRD; families and natives honoured; redactions verified by sampling; manifest hashed and exchanged
- ☐ Cross-cutting: hash continuity demonstrable end to end; data-protection minimisation and disposal terms in place; certificate signed against the checkpoint documents

Red flags

- An opponent who cannot produce the stage deliverables, no data map, no hit reports, no exception log, behind a signed certificate.
- Volumes that do not reconcile between stages, or reductions with no logged basis.
- "We searched everything" with no methodology, or search terms agreed without a single hit report.
- Chat evidence produced only as screenshots or PDFs, with the native database unmentioned.
- Structured systems disclosed as raw dumps, or not addressed at all.
- Review begun, on either side, before the protocol and issues list existed.
- A production with broken families, live text under redactions, or spreadsheets flattened to PDF.
- Your own provider unable to show you a suppression log on request, today.

§ 17 · ESCALATION & COMMON QUESTIONS

When to involve a digital forensic expert

The model hands off to digital forensics whenever the question shifts from *what exists* to *what happened*: authenticity challenges (the test report with the impossible date), suspected deletion or exfiltration, contested attribution, locked or damaged sources, and any assertion about device activity that will need CPR Part 35 / CrimPR Part 19 opinion evidence. The practical rules: preserve forensically at collection wherever such questions are plausible (the image supports every later strategy; an export forecloses some), build the escalation paragraph into the review protocol, and instruct examination by written questions, not open curiosity. Our companion guide *eDisclosure or Digital Forensics? Choosing the Correct Approach* covers the triage in full.

Frequently asked questions

What actually happens to evidence between collection and production?

It is expanded, indexed, stripped of system files, de-duplicated, threaded, OCR'd, filtered by date and negotiated terms, prioritised by analytics, read (in part) by lawyers, privilege-checked, redacted where required, and assembled into an ID'd, hash-manifested production, with every excluded item suppressed rather than deleted, every suppression logged and reversible, and volumes reported at each step. In *Halworth v Brydon* that pipeline turned 2.14 million collected items into 9,650 produced documents without a single item being destroyed, and §§7–11 of this guide walk each step with the numbers.

Is the EDRM binding in England & Wales?

No, it is an industry reference model, not law. The binding framework is PD 57AD or CPR 31/PD 31B (and CPIA in criminal matters). But the model is how methodology is described, quoted and negotiated in practice, and the DRD effectively asks you to document your EDRM, so fluency in it is fluency in the procedure.

Do small cases need all seven stages?

Every case passes through all seven, the question is weight, not existence. A two-custodian tribunal claim still needs a hold, a documented collection, some culling, a review and an orderly production; each stage may be a page of notes rather than a workstream, but skipping one entirely is where small cases go wrong.

Why was only a fraction of the population ever read by a lawyer?

Because reading all of it is neither affordable nor more accurate. The unread remainder is not ignored: it is term-negative or ranked low by validated analytics, sampled statistically, and retained intact. Courts have accepted this architecture since *Pyrrho* (2016); the defensibility lives in the sampling numbers, which is why the checkpoint reports matter.

Who owns each stage, us or the provider?

Identification and preservation: lawyer-led with client IT. Collection and processing: provider-executed under written lawyer instruction. Review: lawyer-owned. Analysis: shared. Production: provider-executed, lawyer-certified. The certificate signatory owns the whole chain, which is why the per-stage deliverables should land on the lawyer's file, not just the provider's.

What happens to the data after the case?

Disposal, the model's implicit final stage: return or verified destruction of collections, review sets and productions under the provider contract and the client's retention policy, subject to appeal windows and any continuing obligations, with destruction certificates on the file. Agree the terms at instruction, not at archive time.

§ 18 · REFERENCE

Glossary

CAL
Continuous active learning, TAR that retrains on every coding decision, feeding reviewers the likeliest-relevant documents first.

Checkpoint deliverable
The document each EDRM phase produces to evidence what was done: data map, hold register, hash manifest, processing report, QC statistics, production manifest.

Data map
The identification-stage record of sources, custodians, locations, volumes, volatility and scoping decisions.

De-NISTing
Removing known system files from a processed dataset by reference to the NIST hash library.

DRD
Disclosure Review Document, PD 57AD's joint record of issues, models, sources and methodology.

EDRM
Electronic Discovery Reference Model, the standard phase map from information governance to presentation.

Exception report
The processing log of items the tools could not open, encrypted, corrupt, unsupported, each resolved or accepted with reasons.

Family
A parent document and its attachments, kept together through review and production.

Hash manifest
The list of cryptographic fingerprints proving a collection or production is complete and unaltered.

Hit report
Per-term counts showing what a search term captures, exchanged and iterated before terms are agreed.

Inclusive message
The email in a thread containing all earlier content, reviewed in place of its predecessors.

Load file
The structured index accompanying a production, mapping documents to metadata, text and images.

Richness curve
The falling rate at which a ranked review stream yields relevant documents, the statistical basis for stopping.

Suppression
Reversible, logged exclusion of an item from the active population, the funnel's mechanism, as distinct from deletion.

TAR
Technology-assisted review, machine-learning classification of documents by likely relevance.

Sources and authoritative references

Stage practice in this guide draws on the EDRM phase notes published by our e-discovery practice, referenced here as sources:

- e-discovery.uk, EDRM Phase 01 · Identification: e-discovery.uk/edrm/identification
- e-discovery.uk, EDRM Phase 02 · Preservation: e-discovery.uk/edrm/preservation
- e-discovery.uk, EDRM Phase 03 · Collection: e-discovery.uk/edrm/collection
- e-discovery.uk, EDRM Phase 04 · Processing ("cull hard, before a lawyer reads the first document"; family-aware deduplication; reversible suppressions; hit and exception reporting): e-discovery.uk/edrm/processing
- e-discovery.uk, EDRM Phase 05 · Review: e-discovery.uk/edrm/review
- e-discovery.uk, EDRM Phase 06 · Production: e-discovery.uk/edrm/production
- e-discovery.uk, Knowledge Centre (practice notes on search-term negotiation, mobile and chat evidence, regulator timescales): e-discovery.uk/knowledge
- EDRM, the reference model and phase guides: edrm.net/edrm-model

- Practice Direction 57AD and the Disclosure Review Document: justice.gov.uk, PD 57AD
- CPR Part 31 and PD 31B, Disclosure of Electronic Documents: justice.gov.uk, PD 31B
- CPR Part 35, Experts and Assessors: justice.gov.uk, Part 35
- *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch), judicial approval of predictive coding: bailii.org
- Data Protection Act 2018: legislation.gov.uk/ukpga/2018/12 · Data (Use and Access) Act 2025: legislation.gov.uk/ukpga/2025/21
- ICO, UK GDPR guidance and resources: ico.org.uk

DISCLAIMER

This publication provides general information about electronic disclosure practice in the United Kingdom as at August 2026. *Halworth v Brydon* is a fictional worked example; any resemblance to real parties is coincidental, and all volumes are illustrative. This guide is not legal advice, does not create a professional relationship, and should not be relied upon in place of advice on the facts of a specific matter from a qualified lawyer or a suitably qualified forensic practitioner. Rules, guidance and legislation change; always check the current text. Links were live at the date of publication.

§ HOW A SPECIALIST LABORATORY CAN ASSIST

Working with Computer Forensics Lab

Our e-discovery practice, **e-discovery.uk**, delivers the model this guide describes as a single in-house capability, forensic collection, processing and culling, Relativity-ready hosted review, and court- or regulator-tuned production, with every phase closing on its defensibility checkpoint and one chain-of-custody discipline from first custodian interview to final manifest. Because the practice sits inside a digital forensics laboratory, the escalation path this guide recommends is internal: when analysis surfaces the anomalous document or the suspect device, the same organisation preserves, examines and, where needed, reports under CPR Part 35 / CrimPR Part 19. We work to your platform requirements, cull aggressively before review (you pay for substance, not volume), host UK-resident and encrypted, and advise candidly, including when a phase can safely be done in-house by your own team. Conflicts checks, confidential scoping discussions and fixed-fee estimates are routine.

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