COMPUTER FORENSICS LAB · E-DISCOVERY TEAM

CFL ELECTRONIC EVIDENCE SERIES · EARLY CASE ASSESSMENT

Benefits of Early Case Assessment in eDiscovery

A Practical Guide to Understanding Evidence, Controlling Disclosure Costs and Developing Case Strategy Earlier

> *Understand the evidence before the evidence controls the case.*

How early data intelligence helps legal teams understand risk, control disclosure costs and develop better case strategy. Written for solicitors, barristers, in-house counsel, disclosure lawyers, investigators and litigation-support professionals in England and Wales.

Position stated as at 1 September 2026 · Approximately 45 minutes reading time

PD 57AD · CPR 31 / PD 31B · CPR 1 OVERRIDING OBJECTIVE · UK GDPR · DUAA 2025 · NPCC · ISO 17025 ALIGNED PRACTICE · CFLAB.UK · E-DISCOVERY.UK

§ 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**, opens every matter with assessment of custodians, repositories, volumes and communication patterns, so that collection, processing and review decisions rest on measured information rather than assumptions. Our examiners also produce CPR Part 35 and CrimPR Part 19 compliant expert reports and give evidence in the Business and Property Courts, County Court, Crown Court and employment tribunals.

ESTABLISHED 2007 · LONDON | ISO 17025 ALIGNED PRACTICE | NPCC DIGITAL EVIDENCE PRINCIPLES | EARLY CASE & DATA ASSESSMENT | HOSTED REVIEW & ANALYTICS | FULL CHAIN-OF-CUSTODY DOCUMENTATION

HOW TO READ THIS GUIDE

Superscript numbers refer to the reference list in section 27. Figures described as hypothetical or illustrative are invented for teaching purposes and are not benchmarks. This publication provides general information, not legal advice (section 28).

§ CONTENTS

In this guide

01	Executive summary and the ECA Principle	3
02	ECA and EDA: two related disciplines, one central proposition	4
03	The traditional problem: collect everything and review later	6
04	ECA and the disclosure framework in England and Wales	7
05	The strategic benefits of ECA (twelve benefits)	10
06	What modern eDiscovery technology adds to ECA	13
07	The modern ECA workflow: twelve stages in a feedback loop	16
08	The evidence universe	17
09	ECA can reveal what is missing: source architecture and alternative evidence paths	18
10	Case study: Pyrrho Investments v MWB Property (2016)	21
11	Case study: Triumph Controls v Primus International (2018)	22
12	Worked case study: Halden Systems v a departing sales director (hypothetical)	25
13	Cost illustration and the eDiscovery cost funnel	26
14	Questions ECA should answer (lawyer's checklist of 54 questions)	28
15	ECA deliverables and sample templates	29
16	ECA dashboard concept (illustrative data only)	30
17	ECA and privilege	30
18	ECA and UK data protection	31
19	Fifteen ECA mistakes that can become expensive later	33
20	When digital forensics should enter the ECA process	35
21	Practical implementation plan: first 24 hours to the DRD	35
22	Question banks: client, opponent and provider	36
23	Model instruction wording	37
24	Red flags	38
25	Glossary	38
26	Conclusion: early knowledge changes the economics of disclosure; specialist support	40
27	References and bibliography	41
28	Disclaimer and jurisdictional notes	42
A	Appendix: detachable Early Case Assessment Checklist for Litigation Lawyers	43

Executive summary

THE ECA PRINCIPLE

"Do not begin by asking how quickly every potentially relevant byte can be collected. Begin by asking what the legal team needs to know, which evidence could answer those questions, where that evidence is likely to exist, and what proportionate method should be used to preserve, assess and collect it."

Early Case Assessment (ECA) is the structured, early analysis of a dispute: its facts, legal issues, risks, likely liability, key witnesses, evidence requirements, probable cost and possible strategy. Early Data Assessment (EDA) is the early analysis of the electronically stored information itself: custodians, sources, volumes, dates, file types, communications, duplication, concepts, keywords and collection requirements. The EDRM Identification Standards treat the two as the twin components of identification.[10] This guide argues that the principal value of ECA is not a smaller review population. It is knowledge, obtained early enough that strategy, scope, budget and preservation decisions can still be changed.

QUESTION	SHORT ANSWER
Why perform ECA before mass collection and review?	Because the most expensive decisions in disclosure (which custodians, which sources, which date ranges, which method) are otherwise taken blind, and PD 57AD expects parties to discuss and seek to agree searches limited by exactly those parameters "with a view to reducing the burden and cost of the disclosure exercise".[3]
What questions can ECA answer?	Who communicated with whom; when activity clustered; which repositories hold the substance; how much material exists and how much is duplicated; where the hot and adverse documents sit; what appears to be missing; what collection and review are likely to cost.
How can it affect strategy?	Adverse documents found early inform merits advice, pleadings, mediation and settlement while options remain open. Supportive material found early shapes witness preparation and applications.
How does it reduce unnecessary collection and review?	By replacing "collect everything" with measured, documented decisions about custodians, sources and date ranges, and by using deduplication, threading, sampling and analytics that PD 57AD and the DRD expressly contemplate.[3,4] Preservation is not narrowed; collection and review are targeted.
How does it identify important and adverse evidence earlier?	Communication analytics, timeline analytics, concept clustering and targeted sampling of priority custodians surface the documents that decide cases weeks or months before linear review would reach them.
How does it help estimate cost?	Measured source volumes, expansion ratios, duplication rates and hit counts convert budget guesses into formula-based estimates that can be defended in a costs budget and in the DRD.
How does it identify custodians and missing evidence?	Metadata and network analysis reveal who actually communicated, including people absent from the organisation chart; gaps in date distributions, missing attachments and one-sided threads show where evidence has moved, been deleted or was never collected.
How does it assist proportionality and the DRD?	Section 2 of the DRD asks for data sources, custodians, date ranges, unavailable sources and search and technology proposals.[4] ECA is the work that allows those answers to be accurate rather than aspirational.

When should specialists become involved?	At the preservation stage, and immediately where deletion, exfiltration, personal devices, encryption, disputed timestamps or attribution are in play (section 20). Instructing early is cheaper than repeating a search under order, as Digicel and Triumph Controls illustrate.[6,7]

FIVE POINTS TO TAKE AWAY

- ECA is not a procedural stage mandated by PD 57AD; it is the practice that lets a party meet PD 57AD's expectations of proportionate, cooperative and technology-aware searching.[3]
- Preservation is broad and immediate; collection and review are targeted and staged. Never confuse the three.
- Technology does not make a methodology defensible. Validation, documentation, senior legal oversight and transparency do (Triumph Controls).[6]
- ECA reveals absence as well as presence: what is missing often matters more than what is found.
- Reduction percentages are case-specific. This guide does not claim that ECA guarantees lower costs or better outcomes; it claims that ECA delivers evidence-based intelligence when strategy can still change.

§ 0 2 · FIRST PRINCIPLES

ECA and EDA: two related disciplines, one central proposition

The central proposition of this guide is simple. The principal benefit of Early Case Assessment is not merely reducing the number of documents lawyers must review. Its greater value is allowing legal teams to understand the evidence, risks, relationships, likely disclosure burden and potential strengths and weaknesses of a matter sufficiently early to make better strategic decisions.

The problem explained in plain English

Most disputes now begin with a paradox. The client holds more potentially relevant information than at any time in history, spread across mailboxes, Teams channels, OneDrive, SharePoint, phones, messaging apps, cloud applications and backups, yet the legal team knows less about that information at the moment it must make its most consequential decisions. Which custodians? Which systems? Which dates? What will it cost? Decisions taken at that point tend to be over-inclusive because over-inclusion feels safe. It rarely is. It multiplies processing, hosting, privilege review and personal-data exposure, and it delays the moment at which the lawyer actually understands the case.

Conventional or poorly implemented ECA has frequently followed a "collect first, understand later" model: identify a large number of custodians; collect extensive datasets; process everything; apply broad keyword searches; send large populations for linear human review; and discover duplication, irrelevance and incorrect assumptions only after substantial expenditure has occurred. The alternative is a data-informed methodology in which the team progressively understands the information before committing to collection and review at scale.

Important terminology

TERM	MEANING IN THIS GUIDE
Early Case Assessment (ECA)	Early analysis of the dispute: facts, legal issues, risks, potential liability, key witnesses, evidence requirements, likely costs and possible strategy. A legal exercise informed by data.
Early Data Assessment (EDA)	Early analysis of the electronically stored information (ESI) itself: custodians, sources, volumes, date distributions, file types, communication patterns, duplication, concepts, keyword performance, potentially relevant material and collection requirements. A technical exercise directed by legal questions.

Preservation	Steps taken to prevent documents being lost or altered once litigation is contemplated. PD 57AD para 3.1 imposes the duty to take reasonable steps to preserve; para 4 sets out what those steps include.[3] Preservation is broad; it is not collection.
Collection	The forensically sound copying of preserved data from its source for processing. Collection can be staged and targeted without narrowing preservation.
Processing	Extraction of text and metadata, expansion of containers and attachments, deduplication and indexing to make data searchable.
Issues for Disclosure	The key issues in dispute that the parties consider will need to be determined by reference to contemporaneous documents (PD 57AD para 7).[3]
Disclosure Review Document (DRD)	The joint document under PD 57AD para 10 recording Issues for Disclosure, proposed Models, data sources, custodians, date ranges, search and technology proposals and cost estimates.[3,4]
Technology-assisted review (TAR)	Review in which software learns from human coding decisions to score or classify documents for relevance. Continuous active learning (CAL) is a TAR workflow in which the model updates continuously as reviewers code. Predictive coding is the older term for the same family. These are distinct from semantic search and from generative AI.[12,13]
Recall and precision	Recall: the proportion of relevant documents actually found. Precision: the proportion of documents found that are relevant. Both were adopted by the English court in Pyrrho when quoting Moore v Publicis.[5]

How modern eDiscovery combines the two

ECA asks the legal questions: what must be proved, what could defeat us, who knew what and when. EDA answers them empirically from the data and, in doing so, generates new legal questions. The disciplines are iterative. A lawyer who reads a suspicious Teams thread revises the issues; a revised issue prompts a new search; the search reveals a custodian who was never interviewed. Section 7 sets out the feedback loop in detail. The EDRM Identification Standards make the same point in their own terms: documenting the findings during identification is as important as discovering the information.[10]

LAWYER'S QUESTION

Is ECA just a vendor label for a processing platform?

TECHNICAL ANSWER Many platforms market an "ECA module". The module is a tool: it indexes a dataset and offers filtering, threading, deduplication and analytics before full review. ECA as a discipline is the decision-making that the tool supports, and it begins before any data is loaded.

PRACTICAL CONSEQUENCE Buying the module does not deliver ECA. Instructing the right questions, interviewing the right people and documenting the answers does. The tool then accelerates the empirical part.

§ 0 3 · THE PROBLEM

The traditional problem: collect everything and review later

Consider a hypothetical commercial dispute over a terminated supply agreement. The client identifies 20 custodians. Their data sits in Microsoft 365 mailboxes, Teams, SharePoint and OneDrive, on several laptops and mobile phones, in WhatsApp, on network file shares, in an archived email journal and in a handful of cloud applications. The first instruction is: "collect everything for all twenty, back to 2019, so we are covered."

What "collect everything" actually produces

Twenty full mailboxes over a multi-year period, twenty OneDrive accounts, every Teams chat and channel those people touched, full forensic images of laptops and phones, entire departmental SharePoint sites and file-share trees, and a journal archive that already contains most of the mail a second time. Processing expands containers and attachments, so the document count grows well beyond the file count. The same email now exists in the sender's mailbox, each recipient's mailbox, the journal, the laptop OST cache and the phone backup. Nobody yet knows which custodians matter, whether the dispute really spans five years or five months, or that the operative discussions moved to Teams in 2023.

CONSEQUENCE	WHY IT HAPPENS
Unnecessary forensic collection and processing	Devices and accounts are imaged and expanded before anyone has tested whether they hold relevant material.
Data expansion, duplicate documents and duplicate email	Attachments, embedded objects and multi-mailbox copies inflate the population; journaling and caches add further copies.
Irrelevant system data and irrelevant custodians	Whole images bring operating-system files, application caches and personal material; peripheral custodians add volume without substance.
Over-broad date ranges and repeated attachments	Ranges chosen defensively rather than from evidence; the same board pack circulated thirty times.
Excessive hosting and privilege review	Hosting is charged on volume; every extra document that could be privileged must be checked before production.
Increased personal-data processing	Every irrelevant employee email, HR thread and personal message collected is personal data processed without necessity (section 18).
Slower identification of important and adverse documents	Linear review reaches the decisive material in proportion to where it sits in the batch order, which is random.
Incorrect assumptions surviving longer	Nothing tests the "twenty custodians, five years, email" assumption until review is well under way.
Delayed settlement analysis and poor cost forecasting	Merits advice waits for review; budgets are guessed from source volume rather than measured populations.
Increased risk of disclosure disputes	Positions in the DRD are aspirational; when reality diverges, the opponent applies to the court.

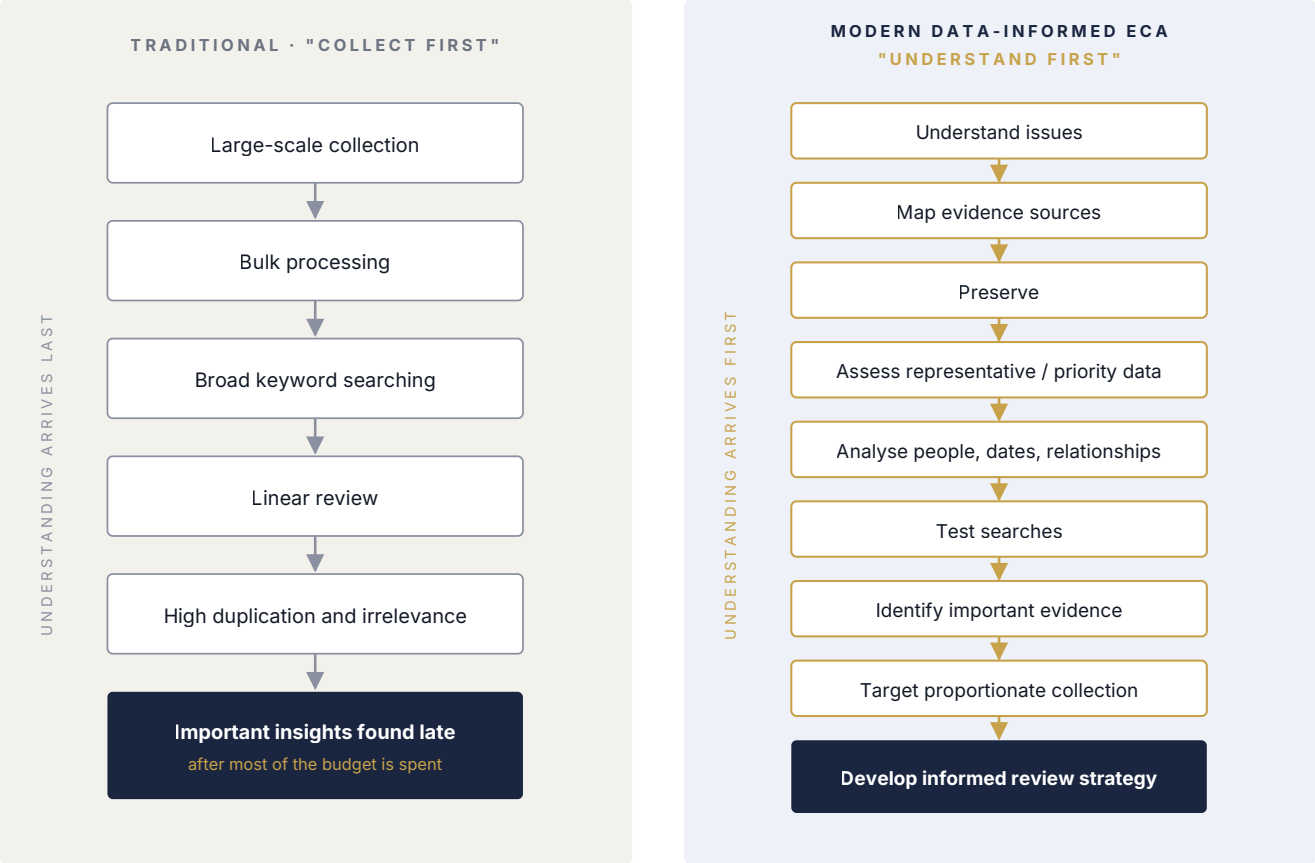


Figure 1 · Traditional ECA versus modern data-informed ECA. In the traditional sequence (Collect, Process, Review Everything, Eventually Understand the Case) insight is the last output. In the modern sequence (Understand Issues, Map Evidence, Assess Data, Test Assumptions, Prioritise, Collect Proportionately, Review Strategically) insight is the first. Alt text: two vertical flowcharts side by side; the left ends in a dark box "important insights found late", the right ends in a dark box "develop informed review strategy".

§ 0 4 · WHY IT MATTERS LEGALLY

ECA and the disclosure framework in England and Wales

ECA is not a procedural stage. No rule or practice direction mandates it by name. What the rules do mandate is proportionality, cooperation, preservation, informed identification of custodians and sources, and a considered approach to technology. ECA is the practical work that makes those obligations achievable and the party's positions credible.

The Business and Property Courts: PD 57AD

Practice Direction 57AD governs disclosure in most Business and Property Courts proceedings (with the exceptions listed at para 1.4).[3] Its architecture maps closely onto what ECA delivers:

- **Preservation (paras 3.1 and 4).** A person who knows it is or may become a party is under a duty to take reasonable steps to preserve documents in its control that may be relevant, including suspending deletion routines and notifying former employees and third parties where appropriate.[3] ECA's preservation-risk review is how that duty is discharged intelligently.
- **Issues for Disclosure (para 7).** The parties must identify the key issues that will need to be determined by reference to contemporaneous documents. Early understanding of the factual story determines whether those issues are drawn well.

- **Disclosure Models (para 8).** Models A to E range from known adverse documents to wide search-based disclosure. Choosing and arguing for a Model requires knowledge of what a search of each source would actually involve.
- **Reasonable and proportionate search (paras 6.4 and 9.6).** Where a Model requires searches, the parties must discuss and seek to agree limits by date range, custodian, class of document, repository, system or device, and automated searches, "with a view to reducing the burden and cost of the disclosure exercise".[3] Every one of those parameters is an EDA output.
- **Technology (paras 9.6 and the DRD explanatory notes).** The parties must consider software and analytical tools, including technology-assisted review, deduplication and sampling, where proportionate.[3,4]
- **The Disclosure Review Document (para 10).** Section 2 of the DRD asks each party to identify data sources and locations, custodians and date ranges, unavailable or inaccessible sources, initial search proposals and proposals to use technology, together with cost estimates.[4] A DRD completed without ECA is a DRD completed on hope.
- **Cooperation and Disclosure Guidance (paras 2.3, 11).** Parties are expected to cooperate and may seek Disclosure Guidance. Measured data supports more constructive discussions with the opponent and the court.
- **Staged disclosure and production planning.** The court may order disclosure in stages or by reference to particular issues; understanding where evidence sits allows a party to propose a sensible sequence.

CPR Part 31 and PD 31B

Outside PD 57AD (for example in the County Court, in the King's Bench Division outside the Business and Property Courts, and in the excluded categories at PD 57AD para 1.4), CPR Part 31 continues to apply, with PD 31B governing electronic documents in multi-track cases.[2] The "reasonable search" factors in CPR 31.7(2) (number of documents, nature and complexity, ease and expense of retrieval, significance of documents likely to be located) and PD 31B paras 20 to 27 on keyword and automated searching were the framework applied in both Pyrrho and Triumph Controls.[5,6] PD 31B's emphasis on early discussion of the scope of searches, and its warning that a party disclosing without prior discussion may be required to repeat searches (para 19, quoted in Triumph Controls at [15]), is a direct procedural argument for doing ECA early and sharing its results.[6]

ECA findings and their possible disclosure consequences

ECA FINDING	POSSIBLE DISCLOSURE CONSEQUENCE
Custodian has little relevant communication	Reconsider collection scope for that custodian; record the reasoning in the DRD.
Relevant period appears limited to six months	Propose narrowed date ranges under para 9.6(1)(a), with the date-distribution evidence.
Most communications involve three people	Prioritise those custodians for early collection and review; stage the rest.
Documents are heavily duplicated	Apply defensible, documented deduplication (section 5.7).
Relevant information sits in Teams rather than email	Redirect collection to Teams and its SharePoint/OneDrive storage; revise DRD section 2 sources.
Significant adverse material identified	Escalate legal assessment: merits, Model A "known adverse documents", settlement posture.
Very large review population	Consider analytics and TAR; prepare the para 9.6 and DRD technology proposal.
Important documents exist in a SaaS platform	Preserve and collect that source; identify export limitations in DRD section 2, question 3.
Critical period predates the current system	Investigate backups, archives and migration remnants; record accessibility issues.

Relevant mobile messaging discovered	Extend the evidence-source assessment to handsets, backups and counterpart devices; consider forensic collection.

JURISDICTIONAL NOTES

Scotland: recovery of documents proceeds by commission and diligence and specification of documents; there is no equivalent of the DRD, but proportionality and specification discipline reward the same early understanding. **Northern Ireland:** discovery under the Rules of the Court of Judicature (NI) 1980, Order 24, follows a broader relevance test; ECA still governs cost and preservation. **Arbitration:** document production is governed by the tribunal's procedural orders and often the IBA Rules on the Taking of Evidence; early data intelligence supports narrow, specific Redfern requests. **Tribunals:** employment tribunals apply proportionality under their own rules, with no formal search-based Models. **Criminal proceedings:** CPIA 1996 and the Attorney General's Guidelines on Disclosure govern; the equivalent discipline is early identification of reasonable lines of enquiry and digital material strategy, and expert reports must comply with the Forensic Science Regulator's statutory Code.[22,23]

§ 0 5 · STRATEGIC BENEFITS

The strategic benefits of ECA

Figure 2 · The ECA intelligence wheel. Twelve categories of intelligence radiate from Early Case Assessment and converge on strategy. Alt text: a central dark circle labelled Early Case Assessment surrounded by twelve pill-shaped labels connected by spokes, with a gold bar labelled Strategy beneath.

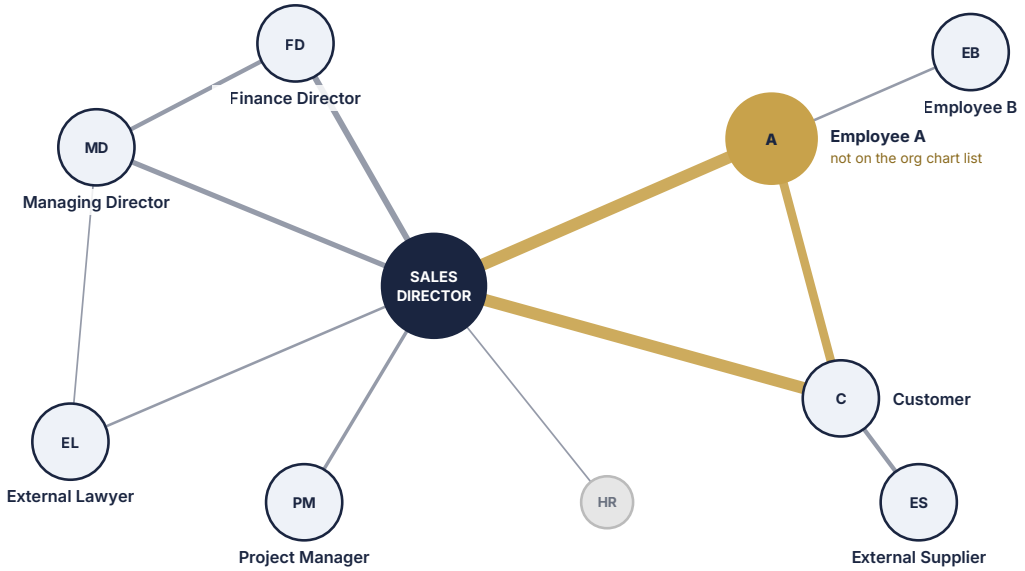
5.1 Understanding the factual story earlier

Chronology is the spine of most commercial cases. ECA reconstructs who was involved, what happened, when, what communications occurred, who knew what, who communicated with whom, which documents changed and what actions followed important communications. Timeline analytics plot document and message activity by day and by custodian; version histories show when a contract draft acquired its contentious clause; threading shows who was copied in when the tone changed. The lawyer arrives at the first client conference with a draft chronology drawn from the documents rather than from memory.

5.2 Identifying the important custodians

Custodian lists built from an organisation chart identify the obvious custodians. ECA distinguishes several further categories: **key custodians** (those whose communications carry the substance); **secondary custodians** (recipients and observers, useful for corroboration and for recovering material a key custodian deleted); **departed employees** (whose mailboxes may be inactive, on litigation hold, or already purged); **shared accounts** (sales@, projects@, service desks) which no individual "owns"; **organisational repositories** (SharePoint sites, file shares, CRM, ERP) that are non-custodial; and **non-custodial data sources** generally (audit logs, VPN logs, badge data). Communication analytics routinely surface a project coordinator or executive assistant who appears in the majority of relevant threads yet was never interviewed. Guide 4 in this series, Finding the Evidence, sets out the custodian and source questionnaire; this guide concentrates on selection and prioritisation.

HYPOTHETICAL COMMUNICATION NETWORK · ILLUSTRATIVE ONLY



Line thickness = relative message frequency in the assessed period. The organisation chart placed Employee A two levels below the Sales Director; the data places Employee A at the centre of the customer traffic.

Figure 3 · Who matters? A hypothetical communication network. Line thickness indicates communication frequency. A person who appeared peripheral on the organisation chart (Employee A) becomes central once message traffic is analysed. Alt text: network diagram of nine labelled nodes; the thickest gold lines join the Sales Director, Employee A and the Customer.

5.3 Finding important documents earlier

The techniques are individually simple and collectively powerful: **date filtering** to the periods where activity clusters; **sender and recipient filtering** to the key relationships; **communication analysis** to reveal the channels actually used; **concept searching** for documents about a subject regardless of vocabulary; **keyword testing** to measure hit counts and false positives before terms are agreed; **clustering** to group similar documents so a reviewer can dismiss a whole cluster of newsletters or read a whole cluster of pricing discussions; **email threading** to isolate the inclusive messages; **metadata analysis** of authorship, editing and transmission; **document similarity** to find near-duplicates of a known hot document; **targeted sampling** of priority custodians; and **chronology analysis** to align documents with the pleaded events. Each is explained, with its limitations, in section 6.

5.4 Discovering adverse documents early

Finding a harmful document early is frequently more valuable than finding a helpful one. A helpful document improves an already viable case; a harmful one may change whether the case should be fought at all. Early discovery of adverse material informs **merits analysis** (prospects advice can be given on evidence rather than instructions), **pleadings** (allegations are not made that a document will later contradict), **settlement and mediation** (offers are calibrated to real exposure while a without-prejudice conversation remains possible), **witness interviews** (witnesses are shown the document before, not after, their statement is drafted), **disclosure strategy** (Model A known adverse documents must be disclosed in any event under PD 57AD, and the duty to disclose known adverse documents continues), **expert evidence** (experts are instructed on accurate facts) and **commercial decision-making** (the board can reserve, provision or exit). Adverse documents do not become less adverse by being found late; they only become more expensive.

5.5 Improving case valuation and risk analysis

Evidential intelligence informs prospects of success, litigation exposure, potential damages, settlement range, cost-benefit analysis and commercial strategy. The distinction that must be kept clear is between technological analysis and legal judgment. Analytics can report that 71 per cent of the pricing discussions occurred between three people in a four-month window; only a lawyer can say what that means for the breach allegation. ECA outputs are inputs to advice, not substitutes for it.

5.6 Targeting collection

ECA can narrow or prioritise custodians, devices, mailboxes, Teams accounts, cloud repositories, date ranges, folders, file types, applications and data sources. Two rules apply. First, narrowing collection must never compromise preservation: everything potentially relevant remains preserved, and only collection and review are targeted. Second, every narrowing decision must be recorded with its evidential basis, because it will be tested by the opponent and, if necessary, the court, as in Digicel and Triumph Controls.[6,7]

5.7 Reducing unnecessary review

Exact deduplication removes byte-identical copies identified by hash value. **Custodial deduplication** removes duplicates within, but not across, custodians, preserving the fact that each custodian held the document. **Near-duplicate identification** groups documents that are textually similar but not identical (drafts, forwards with added text). **Email threading** identifies the inclusive messages that contain all earlier content in a thread. **Exclusion of irrelevant system material** removes operating-system and application files by hash against known-file libraries. **Search-term refinement**, **analytics** and **prioritisation** then order what remains. The distinction to preserve is between reducing review burden and improperly excluding potentially disclosable material: deduplication removes copies, not documents; threading defers non-inclusive messages, it does not delete them; and the suppressed items remain in the dataset, logged and retrievable.

5.8 Better budgeting

Measured ECA outputs (source volume in gigabytes, expansion ratio after processing, deduplication rate, hit rates for tested terms, family sizes, proportion of documents from legal domains) improve estimates for forensic collection, source-data volume, processed volume, hosting, document review, privilege review, redaction, TAR and analytics, production and expert support. Section 13 shows the arithmetic. In PD 57AD proceedings the DRD asks for disclosure cost estimates, and CPR Part 3 costs budgeting requires them; ECA is how those figures acquire a basis.[1,4]

5.9 Better settlement decisions

Parties who understand their evidence earlier are better placed to evaluate settlement, mediation, discontinuance, narrowing of issues, admissions and targeted applications. This guide does not claim that ECA improves outcomes; a bad case remains a bad case. The claim is narrower and more defensible: decisions taken with knowledge are better-informed than decisions taken without it, and the cost of acquiring that knowledge early is usually a small fraction of the cost of acquiring it through full review.

5.10 Better disclosure negotiations

A party that can say "the three proposed additional custodians generated 1,900 documents in the relevant period, of which a 200-document sample produced two marginally relevant items" is negotiating from data. A party that can only say "we think those custodians are peripheral" is negotiating from assumption, and PD 31B para 19 and Digicel show what happens to unilateral assumptions.[2,7] Factual information about volumes and repositories also allows the parties to explain to the court, at a Disclosure Guidance hearing or CMC, why a proposed scope is proportionate.

5.11 Better preservation decisions

ECA routinely reveals disappearing cloud data (Teams chat retention set to 30 days), short retention periods, auto-delete policies, departing employees whose accounts will be purged on a schedule, expiring backups, mobile applications with ephemeral messaging, SaaS systems with limited export, and overlooked archives. Each is a preservation risk under PD 57AD para 4 and each is cheaper to address in week one than to explain in a witness statement about lost documents.[3]

5.12 Better data-protection outcomes

Targeted, proportionate collection reduces unnecessary processing of employee information, personal communications, irrelevant personal data, special-category information and third-party data. UK GDPR Article

5(1)(c) requires personal data to be adequate, relevant and limited to what is necessary; establishing, exercising or defending legal claims is a recognised basis for processing special-category data under Article 9(2)(f), and the DPA 2018 provides exemptions where disclosure is required by court order or is necessary for legal proceedings.[15,16,18] None of this permits a party to avoid disclosure by citing data protection; it means that collecting everything "just in case" is not merely expensive but potentially difficult to justify (section 18).

§ 0 6 · TECHNOLOGY

What modern eDiscovery technology adds to Early Case Assessment

Each technique below is introduced by answering three questions: what is it, why should a lawyer care, and what can go wrong. Human validation remains necessary in every case, because each tool answers a narrow question well and a broad question badly.

TOOL	WHAT IT DOES	WHAT QUESTION IT ANSWERS	KEY LIMITATION
Metadata analysis	Reads dates, authors, custodians, recipients, domains, file types and other attributes without opening documents.	Who created what, when, and who received it? Which file types and domains dominate?	Metadata can be altered by careless collection or by users; dates carry time-zone assumptions; absence of metadata is not absence of activity.
Deduplication	Identifies exact duplicates by cryptographic hash (global or within custodian).	How many unique documents are there? Which copies can be suppressed from review?	Different hash scopes give different counts; emails with trivial header differences are not "exact" duplicates; suppression must be logged.
Near-duplicate analysis	Groups documents that are substantially similar but not identical.	Which documents are drafts or variants of a known document? Which cluster can be reviewed as a group?	Similarity thresholds are configurable; a one-line change in a contract draft may be the decisive change and must not be treated as a duplicate.
Email threading	Reconstructs conversations and flags the inclusive messages containing all earlier content.	What is the full conversation? Which messages need not be read separately?	Threading depends on headers; forwarded or edited quotes, and messages in other systems, break the chain. Attachments differ between messages in a thread.
Communication / network analysis	Maps who communicated with whom, how often and through which domains.	Who are the central participants? Which relationships are unexpected? Which external domains matter?	Measures frequency, not importance; a single decisive message from a quiet participant will not appear central.
Timeline analytics	Plots document and message volume over time by custodian, source or term.	Where are the peaks, gaps and significant periods?	Gaps may reflect collection scope or retention, not events; sent dates, received dates and last-modified dates answer different questions.
Keyword analysis	Runs proposed terms against the indexed set and reports hits, unique hits and families.	Do the proposed terms work? Which are over- or under-inclusive?	Hit counts do not measure recall; a term that returns 40,000 hits is not "working"; misspellings, abbreviations and code words evade terms.

Conceptual analytics	Identifies related documents where the same vocabulary was not used, by statistical modelling of term co-occurrence.	What else is about this subject? What are the main themes in the data?	Works poorly on very short messages, images and spreadsheets; themes require human labelling; not a substitute for search.
Clustering	Groups documents by textual similarity into browsable clusters.	What kinds of documents are in this population? Which clusters can be dismissed or prioritised?	Clusters are statistical, not legal, categories; a cluster labelled "invoices" may contain the one invoice that matters.
Sampling	Draws random or stratified samples so that assumptions can be tested against representative data.	What proportion of this population is relevant? Is the exclusion of a source safe?	Sample size, confidence level, margin of error and the process must be recorded; an undocumented 1 per cent sample was found inadequate in Triumph Controls.[6]
TAR / continuous active learning	Learns relevance from human coding and scores or ranks the remaining population; CAL updates continuously as review proceeds.	Which documents are most likely relevant? When can review stop with measurable recall?	Requires consistent senior coding decisions (Pyrrho at [20]) and validated recall measurement; performs poorly on non-text material; must be disclosed and explained.[5,13]
Generative AI	Large language models that summarise, classify, extract and answer natural-language questions over documents.	What is this set about? Which documents mention X? Draft a chronology from these items.	Hallucination, inconsistency, confidentiality, reproducibility and audit-trail risk; outputs require human verification and professional accountability.[12,14]

Figure 4 · The modern ECA technology toolbox. Eleven techniques, each with what it does, the question it answers and its key limitation. Alt text: four-column matrix of eDiscovery analytics tools.

Generative AI in ECA: emerging uses and necessary safeguards

Generative AI is being used, cautiously, for summarisation of large sets, preliminary issue identification, chronology assistance, relevance assistance, query generation, privilege flagging and interactive interrogation of documents. The Law Society's practical guide on generative AI in legal disclosure (November 2025) and ILTA's Generative AI Best Practice Guide (September 2025) both promote clarity, accuracy checks and human-led review, and treat generative AI as an aid to, not a replacement for, legal judgment and defensible review methodology.[12,13] The Divisional Court in Ayinde v London Borough of Haringey confirmed that lawyers remain fully accountable for AI-generated content and may face regulatory consequences if they fail to verify it.[14]

BEFORE GENERATIVE AI TOUCHES ASSESSMENT DATA, CONFIRM

- **Confidentiality and data security:** where is the data processed, is it used to train models, and is the vendor contract adequate for privileged and personal data?
- **Hallucination controls:** every summary, extracted fact or citation is traced back to the source document and checked by a human before it is relied on.
- **Reproducibility and audit trail:** prompts, model versions, settings and outputs are logged so the exercise can be explained and repeated.
- **Privilege:** AI-generated summaries of privileged material are themselves privileged work product only if created and handled as such; do not put them in the disclosure database.
- **Defensibility:** if the tool influences what is or is not reviewed, be prepared to explain that in the DRD and to the opponent, as with any technology under PD 57AD para 9.6.[3,12]

LAWYER'S QUESTION

Are predictive coding, TAR, CAL, semantic search and generative AI the same thing?

TECHNICAL ANSWER No. Predictive coding and TAR describe supervised classification trained on human relevance decisions; CAL is a TAR workflow that retrains continuously. Semantic or concept search retrieves documents by meaning without a trained relevance model. Generative AI produces new text (summaries, answers, classifications with explanations) from a language model and is not, by itself, a validated relevance classifier.

PRACTICAL CONSEQUENCE Use precise language in the DRD and in correspondence. "We will use TAR" and "we will use generative AI to assist review" are different proposals with different validation requirements.[12,13]

§ 0 7 · PRACTICAL WORKFLOW

The modern ECA workflow: twelve stages in a feedback loop

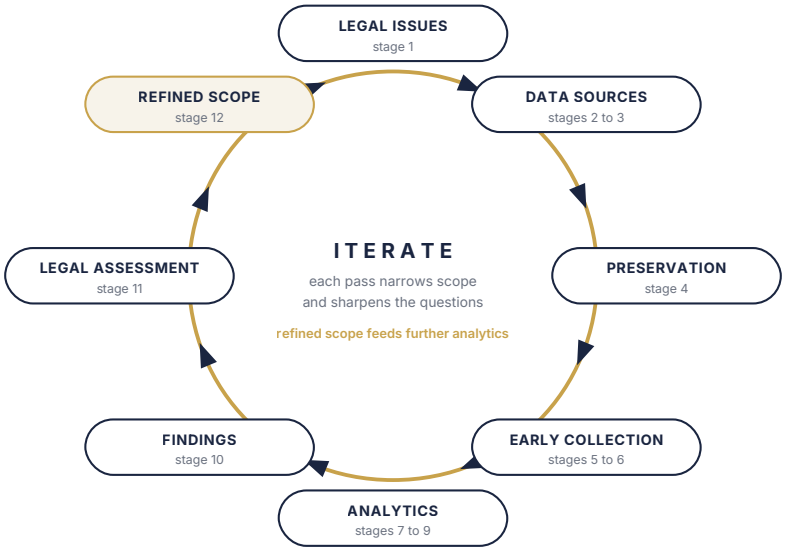


Figure 5 · The ECA feedback loop. Legal issues, data sources, preservation, early collection, analytics, findings, legal assessment and refined scope form a cycle; refined scope feeds further analytics. ECA is iterative, not a single one-off search. Alt text: eight labelled pills arranged around a gold circle with directional arrows; the centre reads Iterate.

STAGE 1 · UNDERSTAND THE DISPUTE

Review pleadings or pre-action correspondence, allegations, chronology, contractual relationships, key individuals and the likely factual questions. Write down the ten things the team most needs to know.

STAGE 2 · CUSTODIAN INTERVIEWS

For each obvious custodian: role, devices, communication channels, applications, storage habits, aliases and shared accounts, historical accounts and system migrations, departed staff they worked with. Ask "where would I look if I were you?"

STAGE 3 · EVIDENCE-SOURCE MAP

Map potential evidence across Exchange and Microsoft 365, Teams, OneDrive, SharePoint, Slack, WhatsApp, laptops, Macs, iPhones, Android devices, network shares, SaaS systems, CRM, ERP, databases, backups, archives, CCTV, logs and third parties (section 8).

STAGE 4 · PRESERVE VOLATILE SOURCES

Identify what could disappear: chat retention, auto-delete, leaver processes, backup rotation, ephemeral messaging, SaaS export windows. Apply holds, suspend routines, forensically image devices where justified, and record every step.

STAGE 5 · ACQUIRE PRELIMINARY DATA

Targeted or staged collection where technically and legally appropriate: priority custodians first, in-place searches where the platform allows, forensic imaging where attribution or deletion may be in issue.

STAGE 6 · CREATE AN ECA DATASET

Process enough to understand the landscape: extract metadata, deduplicate, thread, index. Record the processing settings (deduplication scope, time zone, exceptions) because they will be asked about.

STAGE 7 · ANALYSE METADATA AND COMMUNICATIONS

Volume by custodian and source, date distribution, file types, domain analysis, communication map, concept clusters. Identify patterns and anomalies.

STAGE 8 · TEST SEARCHES

Evaluate recall and precision indicators, false positives, spelling variants, abbreviations, code words, aliases and project names. Sample the hits and, critically, sample the non-hits.

STAGE 9 · LOCATE HOT DOCUMENTS

Prioritise strategically important material using the analytics and targeted reading of priority custodians in the key period. Escalate adverse findings to the supervising lawyer immediately.

STAGE 10 · IDENTIFY EVIDENCE GAPS

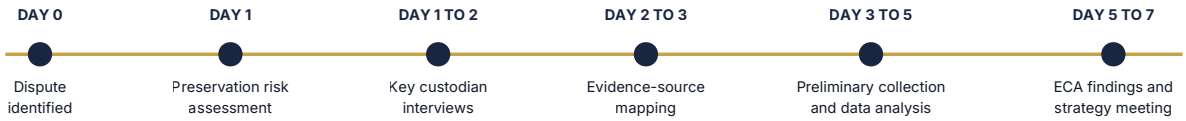
Which expected communications are missing? Unexplained date gaps? Absent attachments? One side of a conversation missing? Incomplete mailbox? Did communication move platform? Historical data in backups? Mobile messages missing? (Section 9.)

STAGE 11 · REASSESS STRATEGY

Feed findings back into merits, pleadings, settlement posture, witness plan and preservation. Revise the Issues for Disclosure if the facts have moved.

STAGE 12 · DEFINE PROPORTIONATE COLLECTION AND REVIEW

Document each decision and its rationale: custodians included and excluded, date ranges, sources, method, technology, sampling. This record becomes DRD section 2 and the answer to any later challenge.



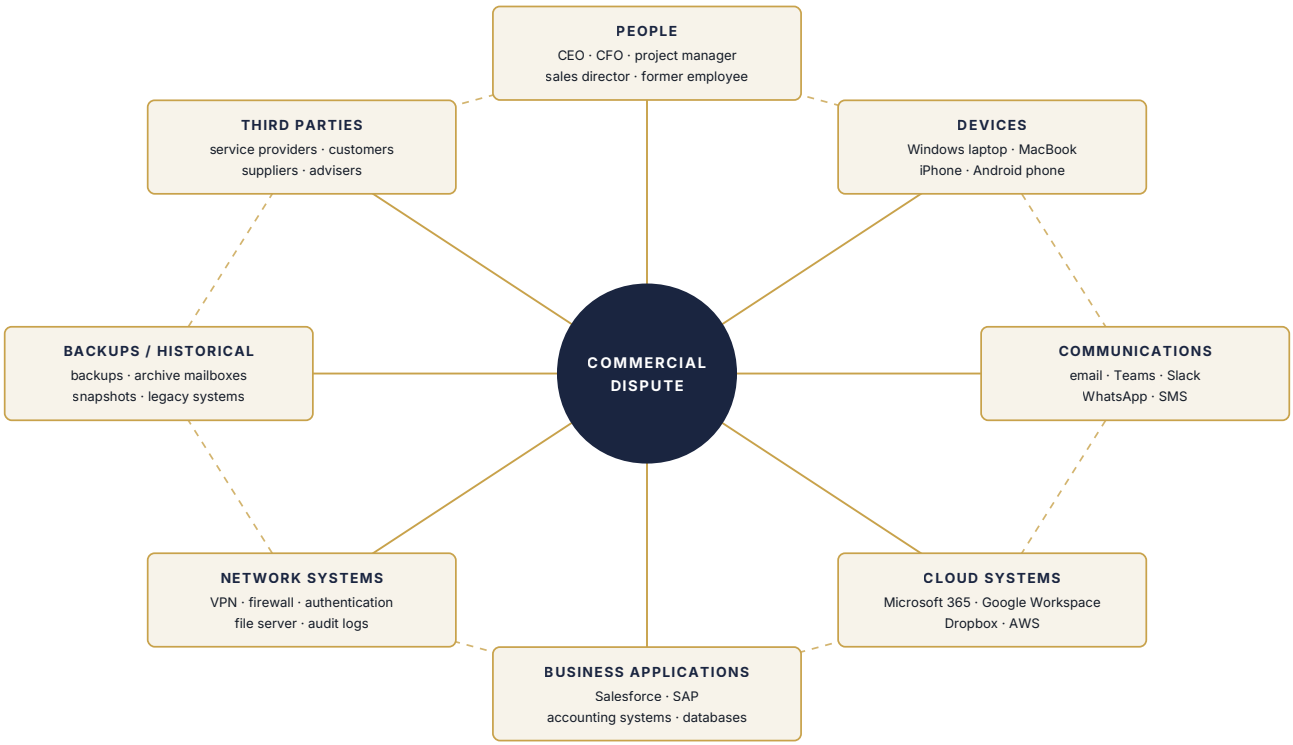
ILLUSTRATIVE WORKFLOW ONLY · NOT A PROCEDURAL DEADLINE · TIMING DEPENDS ON THE MATTER

Figure 6 · A first-week ECA timeline. An illustrative sequence from dispute identification to a strategy meeting within seven days. Alt text: horizontal timeline with six milestones from Day 0 to Day 5 to 7.

§ 0 8 · EVIDENCE-SOURCE MAP

The evidence universe

ECA should identify not merely "documents" but the systems and relationships capable of producing evidence. The evidence universe of a commercial dispute has eight regions, and the same fact usually leaves traces in several of them.



Dashed gold lines: the same evidence commonly exists in more than one region (an email on a device, in the cloud, in a backup and with the counterparty).

Figure 7 · The evidence universe and evidence-source map. A hypothetical matter at the centre, eight regions around it, and cross-links showing that evidence frequently exists in more than one location. Alt text: hub diagram with a central dark circle and eight labelled boxes joined by solid spokes and dashed cross-links.

REGION	WHAT TO ENUMERATE IN ECA	TYPICAL SURPRISES
People	Obvious, key, secondary and departed custodians; assistants; shared accounts.	The assistant who ran the diary is in every meeting invitation.
Devices	Corporate and personal laptops, desktops, phones, tablets, USB media, replaced or reissued devices.	The laptop was reimaged on departure; the phone is personal and used for work.
Communications	Email, Teams (chat and channel), Slack, WhatsApp, Signal, SMS, voicemail, video-call recordings and transcripts.	Substantive negotiation moved to a WhatsApp group in month four.
Cloud systems	Microsoft 365 and Google Workspace tenants, personal cloud accounts, file-sharing services, cloud infrastructure logs.	OneDrive version history shows the document existed before the claimed creation date.
Business applications	CRM, ERP, finance, HR, ticketing, project tools, bespoke databases and their audit trails.	CRM notes record customer calls that no email mentions.
Network systems	VPN, firewall, identity provider, file-server and print logs, badge and door systems.	Sign-in logs place the account in another country on the critical day.
Backups and historical	Backup rotations, archive mailboxes, snapshots, legacy systems, migration exports.	The 2019 mailbox survives only on a tape nobody can read without the retired server.
Third parties	Counterparties, service providers, advisers, regulators, banks, cloud vendors holding logs.	The counterparty holds the only copy of the attachment the client's system stripped.

§ 0 9 · EVIDENCE GAPS

ECA can reveal what is missing

Absence is evidence of something. An abrupt silence, a referenced attachment that is not there, a name that appears in every thread but on no custodian list: each is a trigger for further investigation, and each has a primary source and one or more alternative evidence paths.

WHAT ECA NOTICES	PRIMARY EVIDENCE SOURCE	ALTERNATIVE EVIDENCE PATH
Emails stop abruptly during an important period	Custodian mailbox	Recipient mailboxes; journal archive; backups; mailbox audit log showing deletions; mobile device cache; Teams for the same period.
WhatsApp is mentioned in email but has not been collected	Custodian's handset	Counterpart handsets; iCloud or Google Drive backup of the app; desktop-client cache; exported chat files sent onward; screenshots in mail.
An employee appears in communication analysis but was not listed as a custodian	That employee's mailbox and devices	Existing custodians' copies; group mailbox; Teams channel posts; CRM activity; HR record of role and dates.
Referenced attachments are absent	Original message with attachment	Sender's Sent Items; SharePoint or OneDrive link target and its version history; recipient copies; document-management system; print logs.
A mailbox was deleted	The mailbox	Inactive mailbox or litigation hold copy; backups within retention; journal; recipients; PST exports on devices; audit log of deletion.

A laptop was replaced	The original laptop	OneDrive and known-folder backup; Exchange Online; Teams; recipient mailboxes; central backup; the replacement device's migration data; leaver process records.
An important document exists only as a later version	The original draft	SharePoint/OneDrive version history; email attachments circulated at the time; author's local temp and autosave; recipients' copies; DMS audit trail.
Teams conversations indicate files existed in SharePoint	The SharePoint library	Recycle bins (first and second stage); retention hold library; audit log of file operations; synchronised copies on OneDrive clients; backups.
Audit logs show downloads from a repository that has not been collected	The repository	The downloading device; USB artefacts; personal cloud sync indicators; email forwards; the audit log itself as evidence of the act.

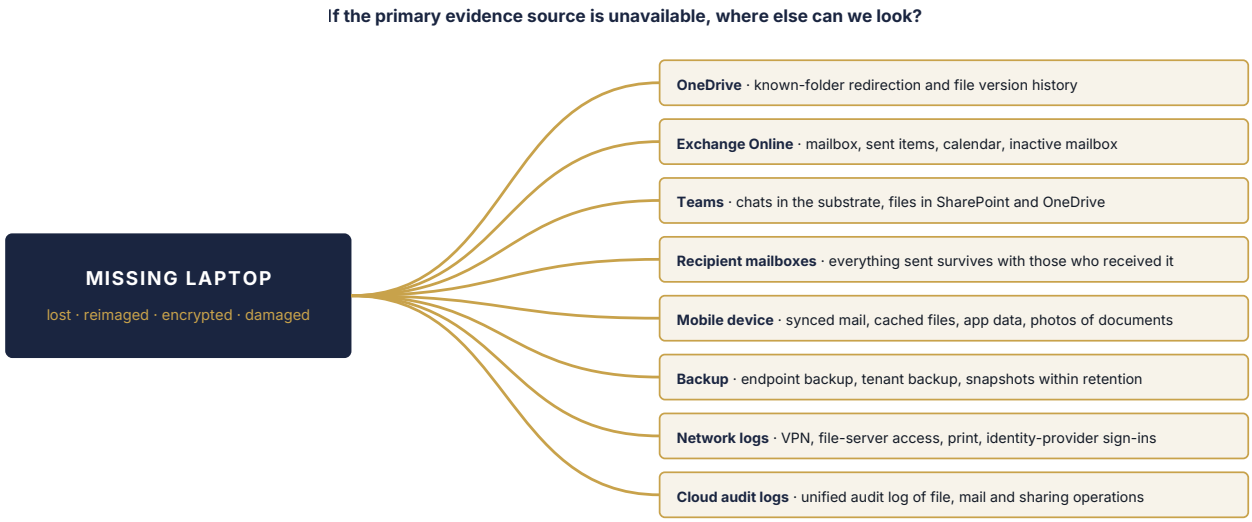


Figure 8 · Missing evidence and alternative paths. A missing laptop branches to eight alternative sources. Alt text: a dark box labelled Missing Laptop on the left with eight curved gold lines to labelled cream boxes on the right.

This is why ECA is as valuable to digital forensic investigators as to disclosure lawyers. The investigator asks the same question in a different register: not "what is disclosable?" but "what happened, and where is the corroboration?" A well-built evidence-source map tells both where to look next when the first source fails.

Source architecture: where else the evidence lives

Do not examine a device or platform in isolation. For an ECA exercise the associated evidence locations are: the physical device; local storage and caches; connected computers that synchronised the same accounts; mobile applications; synchronised accounts; cloud platforms and their version histories; backups; administrative consoles (compliance centre, admin audit); identity providers (sign-in and token logs); network infrastructure (VPN, proxy, firewall); third-party service providers (SaaS vendors, telecoms, cloud hosts); and counterpart devices held by the other participants in a communication.

EVIDENCE	LOCAL DEVICE	CLOUD	BACKUP	LOGS	COUNTERPART DEVICE	DELETED / RECOVERABLE
Email	Y (OST/ PST cache)	Y	Y	Y (mailbox audit)	Y	Often (recoverable items, inactive mailbox)
Teams / Slack messages	Limited (client cache)	Y	Sometimes (tenant backup)	Y (audit log)	Y (same tenant only)	Varies (retention policy)

WhatsApp / mobile messages	Y	Sometimes (app backup)	Sometimes	Limited (metadata only)	Y	Varies (forensic extraction)
Files and documents	Y	Y (with versions)	Y	Activity logs	Possibly (if shared)	Varies (recycle bins, carving)
CRM / ERP records	n/a	Y	Y (vendor)	Y (field audit)	n/a	Sometimes (vendor retention)
Location	Y	Often	Sometimes	Sometimes (sign-in IP)	n/a	Varies
User activity (logins, downloads, USB)	Y (registry, event logs)	Y (unified audit)	n/a	Y	n/a	Varies (log rotation)

Fallback strategy: when the primary source is unavailable, damaged, encrypted, deleted or inaccessible, pursue in order the cloud copy and its version history, the counterpart or recipient copy, the backup within retention, then the logs that prove the document or event existed even if its content cannot be recovered; instruct a forensic examiner before the remaining source is touched.

§ 10 · CASE STUDY

Case study: *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch)

HISTORICAL NOTE

Pyrrho was decided on 16 February 2016 by Master Matthews under CPR Part 31 and PD 31B. It predates the Disclosure Pilot (PD 51U, 2019) and the present PD 57AD regime and must not be presented as a decision applying PD 57AD. Its continuing significance is educational: it is the first English judgment to explain and approve predictive coding, and its reasoning about scale, cost and proportionality remains instructive.[5]

The facts

The claim concerned alleged breaches of fiduciary duty by directors, with a value said to run into the tens of millions of pounds (at [2] to [4]). The bulk of relevant documents sat on back-up tapes holding the email accounts of the second to fifth defendants. The total number of electronic files restored from those tapes was originally more than 17.6 million, reduced to some 3.1 million by electronic de-duplication, which the Master described as "still a large and costly number to search" (at [5]).[5] Electronic Documents Questionnaires had been exchanged and the parties, after several rounds of correspondence, agreed the automated method and the keywords, subject to the court's approval (at [15]).

Why the court approved predictive coding

At [33] the Master listed ten factors: (1) experience in other jurisdictions that predictive coding can be useful; (2) no evidence that it leads to less accurate disclosure than manual review or keyword-plus-manual review, and some evidence to the contrary (citing Moore v Publicis and Irish Bank Resolution Corporation v Quinn); (3) greater consistency in applying a senior lawyer's judgment across the whole set than in using dozens or hundreds of lower-grade fee-earners; (4) nothing in the CPR or Practice Directions prohibits it; (5) the number of documents, over three million, is huge; (6) the cost of manual search would amount to several million pounds at least, so a full manual review would be "unreasonable" within PD 31B para 25; (7) estimated costs of using the software of between £181,988 plus monthly hosting of £15,717 and £469,049 plus monthly hosting of £20,820, depending on factors including whether keyword searches reduced the population; (8) the value of the claims in the tens of millions made those costs proportionate; (9) trial was not until June 2017, leaving time to consider other methods if necessary; and (10) the parties had agreed the use of the software and how to use it. There were no factors of weight the other way, and approval in other cases would depend on their circumstances (at [33] to [34]).[5]

The judgment also records, at [19] to [23], a description of the predictive coding protocol (data set, sample size, control set, confidence level, margin of error), the training sample of 1,600 to 1,800 documents, the best practice that a single senior lawyer who has mastered the issues should consider the whole training sample, and validation by statistical sampling with overturns reviewed by a senior reviewer over usually not fewer than four rounds.[5]

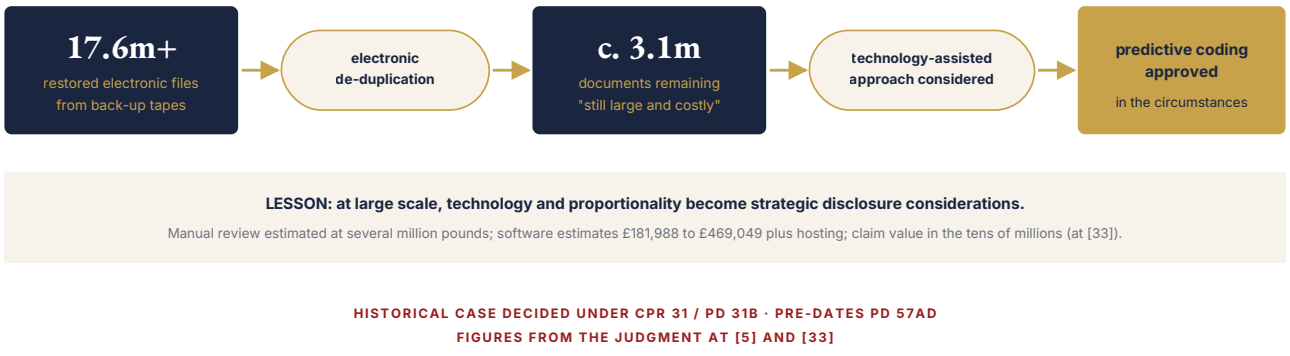


Figure 9 · Pyrrho in numbers. From more than 17.6 million restored files to about 3.1 million after de-duplication, and from there to approval of predictive coding. Source: [2016] EWHC 256 (Ch) at [5], [33].[5] Alt text: five-step horizontal flow from 17.6m files to predictive coding approved.

ECA lessons

1. **Determine scale early.** The parties knew the size of the restored population before arguing about method.
2. **Quantify the data population.** "Over three million" was a measured figure, not an estimate.
3. **Remove unnecessary duplication appropriately.** De-duplication removed roughly five in six files before any lawyer looked at content.
4. **Compare review methods.** Manual review and software were costed side by side.
5. **Calculate proportionality.** Cost was assessed against claim value, as CPR 1.1 and PD 31B require.
6. **Consider technology rather than assuming linear review.** Nothing in the rules prohibited it; the question was suitability.
7. **Agree methodology where possible.** Agreement was the tenth and, in practice, the decisive factor.
8. **Document the process.** Protocol, sample sizes, confidence levels and overturn rounds were all specified in advance.

LESSONS FOR TODAY'S LAWYER

Under PD 57AD the same reasoning appears in para 9.6 and in DRD section 2: measure the population, propose technology where proportionate, agree the approach, and record the protocol. Pyrrho's insistence on a single senior lawyer training the model was later treated by Coulson J in Triumph Controls as best practice rather than a strict rule, but the absence of any such oversight counted against the party that had used the technology (at [30]).[5,6] In Brown v BCA Trading [2016] EWHC 1464 (Ch) the court went a step further and approved predictive coding where the parties did not agree.[24]

§ 11 · CASE STUDY

Case study: *Triumph Controls UK Ltd v Primus International Holding Co* [2018] EWHC 176 (TCC)

Decided by Coulson J on 7 February 2018 in the Technology and Construction Court, under CPR Part 31, PD 31B and the TeCSA/TECBAR eDisclosure Protocol. The judgment demonstrates that technology does not make an eDiscovery methodology automatically defensible.[6]

What happened

The claimants claimed some US$65 million for breach of warranty following the purchase of an aerospace business (at [1]). Their Electronic Documents Questionnaire stated that, following a keyword search, all responsive documents would be "manually reviewed"; it made no reference to computer-assisted review (at

[6]). The agreed keywords, reduced by agreement because of volume, produced 450,000 responsive documents (at [7]). The first list, based on a review of over 200,000 documents, disclosed 12,476 documents; a supplemental list added 4,163 (at [8] to [9]).

It later became apparent that, having reviewed around 230,000 of the 450,000 documents using manual searches aided by CAR, the claimants decided not to search the balance of 220,000 documents that CAR had de-prioritised. They said they had sampled 1 per cent of those documents using a CAR technique, which predicted that only 0.38 per cent would be relevant, and concluded that further searching was disproportionate. This was never discussed with, let alone agreed by, the defendants (at [10]). Ten supplemental lists followed, adding almost 3,000 documents, around 2,000 of which were responsive to the original keywords and so, on the face of it, should have been disclosed originally (at [11]).

A separate complaint about the 860,000 folders on a three-terabyte shared drive failed: the claimants had explained in their list how custodians identified likely folders, only two further folders had since emerged, and the defendants could point to no missing folder (at [19] to [24]).

Why the sampling and review approach was found inadequate

Coulson J's concerns about the 220,000 documents were specific (at [27] to [31]): the claimants did not do what the EDQ said they would do; the disclosure list did not make the method clear; at no time were details provided of how the CAR was set up or operated, which was unsatisfactory where the decision was unilateral; no information was given about how the sampling was conducted, with no stated tolerances and no explanation of the number of rounds; perhaps ten paralegals and four associates were involved with no overseeing senior lawyer performing the role advocated in Pyrrho, so the system may not have been "educated" as well as it might have been; and, in all, neither the CAR nor the sampling could be described as transparent or independently verifiable.

That would not have mattered had the results been reliable. But (at [33] to [36]) the total disclosed (about 19,500, some 4.3 per cent of 450,000 responsive documents) appeared very modest for a document-heavy warranty claim; around 2,000 later-disclosed documents had been keyword-responsive all along; some further documents surfaced only because the claimants' own witnesses wanted to rely on them; and the 0.38 per cent prediction was therefore likely to be an underestimate. The judge concluded that the steps taken were not adequate (at [37]). Because the CAR process had never been explained, "only a manual review will do" (at [38]). Balancing the £180,000 and two-month estimate against a US$65 million claim and a trial less than six months away, he ordered the parties to agree a methodology by which a sample of 25 per cent of the 220,000 documents would be manually searched within three weeks, with results reported in an agreed letter (at [39] to [42]).[6]

> **THE JUDGMENT DOES NOT SAY**
>
> that a 1 per cent sample is always insufficient, that CAR is unreliable, or that manual review is superior. Coulson J expressly accepted that "there is no magic in a manual review" (at [38]). Adequacy depends on context and methodology: whether the process was described in advance, discussed with the opponent, documented with tolerances and rounds, overseen by a senior lawyer, and borne out by the results.[6]

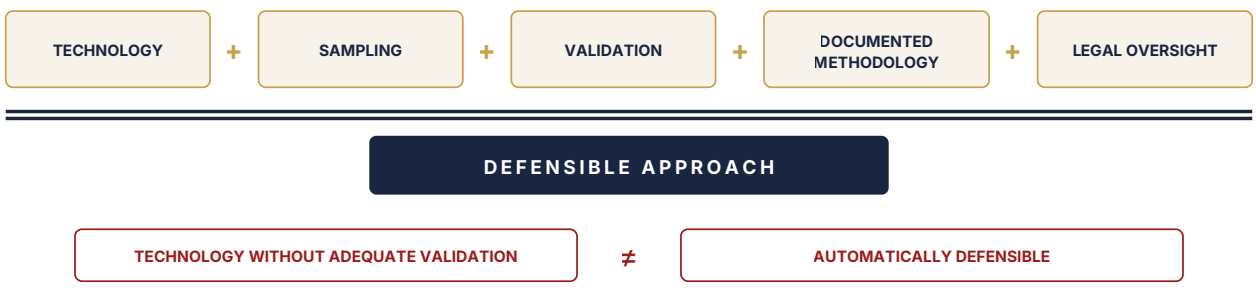


Figure 10 · Methodology must be validated. Technology, sampling, validation, documented methodology and legal oversight together produce a defensible approach; technology without adequate validation is not automatically defensible. Source: Triumph Controls at [27] to [38].[6] Alt text: an equation of five boxes summing to Defensible Approach, with a red "not equal" contrast below.

ECA lessons

- **Documenting methodology:** what was done, by whom, with what settings, must be recorded as it happens; reconstructing it later, as at [29], rarely convinces.
- **Appropriate sampling:** sample size, confidence level, tolerances and rounds are stated in advance; the non-reviewed population is sampled by humans, not only scored by the model.
- **Validating assumptions:** the 0.38 per cent figure was an assumption dressed as a measurement; the later disclosures tested it and it failed.
- **Transparency:** tell the opponent what you propose before you do it (PD 31B para 19; TeCSA protocol paras 4.2 and 5.2; PD 57AD para 9.6).
- **Proportionality:** is measured against the value and nature of the claim, not against the review budget alone.
- **Changing strategy when evidence undermines assumptions:** when witnesses started producing undisclosed documents, the methodology needed revisiting; the court did it instead.

§ 12 · WORKED EXAMPLE

Worked case study: *Halden Systems Ltd* and a departing sales director

HYPOTHETICAL

Halden Systems Ltd, its people, its data and every figure in this section are fictional and used for teaching only. Any resemblance to real persons or companies is coincidental.

Halden Systems, a UK software company, alleges that its former sales director, Mr Voss, diverted major customers to a competitor before resigning. It seeks injunctive relief and damages for breach of contract and fiduciary duty and threatens proceedings in the Business and Property Courts. Potential evidence includes Exchange Online, Teams, OneDrive, the company laptop, an iPhone, WhatsApp, Salesforce, VPN logs, USB activity and indicators of a personal cloud-storage account.

Initial position

The litigation team's first instinct is full collection for twelve custodians (Voss, the sales team of six, the CEO, CFO, head of customer success, two account managers) over two years, on the basis that "we do not yet know what we are looking for". The ECA exercise takes nine working days. Preservation is applied to all twelve custodians and all sources on day one. Collection and analysis begin with three.

INITIAL ASSUMPTION	ECA FINDING	STRATEGIC CONSEQUENCE
Twelve custodians need full collection.	Communication analysis shows that Voss, one account manager and the head of customer success generate 78 per cent of customer-related traffic; several sales-team members appear only in group broadcasts.	Collect and analyse the three first; keep the other nine preserved; stage further collection against findings. Record the basis in the DRD.
The relevant period is two years.	Timeline analytics show customer-contact volume flat until March, then a sharp rise in Voss's external traffic over four months to his resignation.	Propose a four-month primary window with a lookback for contract-formation documents; explain the date distribution to the opponent.
Email is the main source.	Email between Voss and the account manager thins in month two; Teams private chat between them triples in the same weeks.	Redirect collection to Teams (chat substrate plus SharePoint and OneDrive files); adjust DRD section 2 sources.
The customer list in the pleadings is complete.	Salesforce export reveals fourteen accounts reassigned to Voss in the window that appear in no email; six later moved to the competitor.	Amend the pleaded list; preserve and collect CRM audit history; consider whether the reassignment itself is evidence of preparation.
Business communications are on corporate systems.	Three emails refer to "the WhatsApp group"; one account manager confirms a group with two customer contacts and Voss.	Forensic extraction of the account manager's handset with consent; letter to Voss's solicitors regarding preservation of his device; consider a counterpart-device route via the customer.
Nothing was taken.	Unified audit log shows Voss downloaded 1,140 files from the sales SharePoint over two evenings in his final fortnight; the laptop shows a USB device connected on the second evening and a personal cloud client installed.	Escalate to forensic investigation (section 20); preserve the laptop image and logs; consider an application for delivery up and an imaging order.

Everything must be reviewed.	Across the three priority mailboxes 41 per cent of documents are exact duplicates and threading reduces the remainder by a further third.	Apply documented deduplication and threading; plan review on inclusive messages; hold the suppressed copies in the dataset.
All original custodians matter equally.	Four of the original twelve have no direct contact with the affected customers in the window.	Do not collect their content now; keep them preserved; revisit if further findings implicate them.

Preservation versus collection, and collection versus review

Nothing in this exercise permits the other evidence to be destroyed or ignored. All twelve custodians' mailboxes, devices and cloud accounts remain preserved under hold for the life of the matter, because the duty in PD 57AD para 3.1 attaches to documents that may be relevant, not to documents the party has decided to collect.[3] Collection is the subset that is copied for processing; it is staged and expanded as findings justify. Review is a further subset, ordered by priority and reduced by documented deduplication and threading. If the opponent later shows that a peripheral custodian matters, the data is still there and can be collected within days.

LAWYER'S QUESTION

If we only collect three custodians, are we exposed to the criticism made in Triumph Controls?

TECHNICAL ANSWER The criticism there was of a unilateral, undocumented and unexplained narrowing after collection, with results that contradicted it. Here the narrowing is at the collection stage, evidence-based, documented, preserved against, and disclosed in the DRD.

PRACTICAL CONSEQUENCE Put the communication analysis, date distribution and sampling results in DRD section 2 and in correspondence. Invite the opponent to identify any custodian they say is missing. That is the cooperative, transparent approach the court expects.[3,6]

§ 13 · COST ILLUSTRATION

Cost illustration and the eDiscovery cost funnel

The figures below are hypothetical and formula-based. They are not market prices, averages or benchmarks. Their purpose is to show how apparently modest decisions taken early propagate through the funnel and multiply at the review stage.

STAGE	PATH A: COLLECT EVERYTHING (HYPOTHETICAL)	PATH B: ECA-INFORMED (HYPOTHETICAL)
Source data	20 custodians, 5 years, all sources: 900 GB.	Preserve the same 900 GB; collect 6 priority custodians for the evidenced window plus targeted repositories: 210 GB.
Processing expansion	Assume the population expands to 4,000,000 documents after containers and attachments are extracted.	Assume 950,000 documents.
Deduplication	Assume 35 per cent exact duplicates across custodians: 2,600,000 remain.	Assume 35 per cent: 617,500 remain.
Filtering (date, file type, system files, tested terms)	Terms untested; date range broad: assume 2,000,000 potentially relevant.	Tested terms and evidenced date window: assume 210,000 potentially relevant.
Threading and prioritisation	Not applied: review population 2,000,000.	Threading defers 30 per cent to inclusive messages: review population 147,000.
Review hours (assume 50 documents per reviewer hour)	2,000,000 / 50 = 40,000 hours.	147,000 / 50 = 2,940 hours.

Review cost at an assumed £X per hour	40,000 X.	2,940 X. At X = £40 (illustrative), £1.6 million versus £117,600 before hosting, privilege review and QC.
ECA cost	n/a	Assessment processing and analyst time on the 210 GB: assume a low single-digit percentage of Path A's review figure. Case-specific.

The 50 documents per hour rate, the 35 per cent duplication, the expansion ratios and the £40 are placeholders chosen for arithmetic clarity. Real rates vary with document type, issue complexity, reviewer seniority, platform and jurisdiction. Reduction percentages are case-specific; some matters have 10 per cent duplication and a relevant period that really is five years.

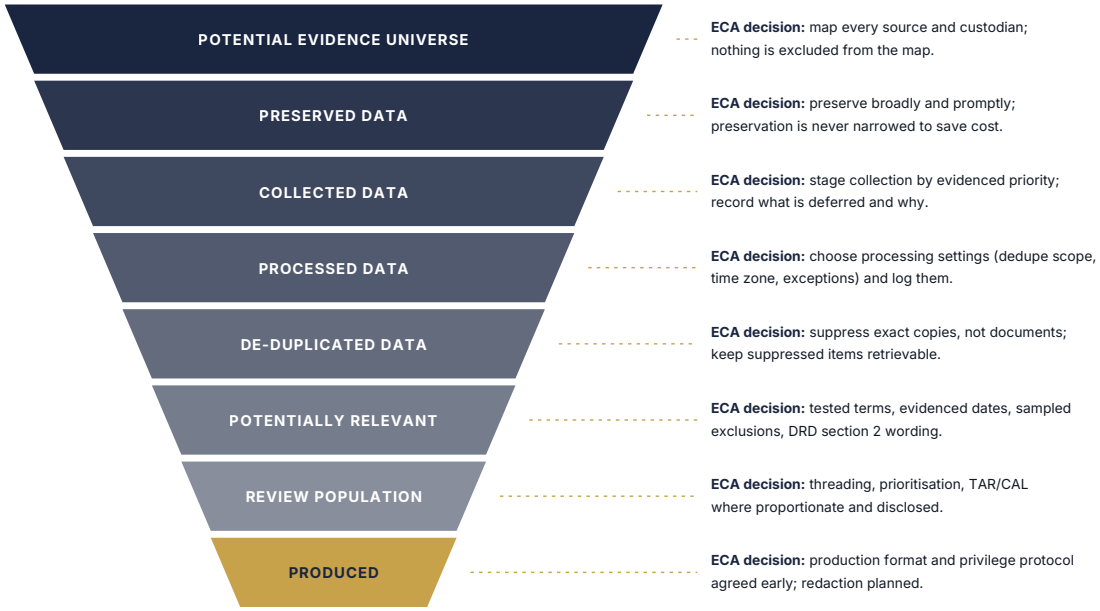


Figure 11 · The eDiscovery cost funnel. Each stage narrows the population; beside each stage is the ECA decision that influences the next. The objective is not indiscriminate reduction: every narrowing decision must be justified and preservation is never reduced to save money. Alt text: eight stacked trapezoids narrowing from Potential Evidence Universe to Produced, with annotations to the right.

§ 14 · LAWYER'S CHECKLIST

Questions ECA should answer

Fifty-four questions organised by theme. If the team cannot answer a question, it has identified a piece of ECA work; if it can answer only by assumption, it has identified a risk.

Case

1. What are the pleaded or anticipated issues, and which will turn on contemporaneous documents?
2. What facts must we prove, and which documents would prove or disprove them?
3. What is the claim worth, and what disclosure spend is proportionate to it?
4. What is the earliest and latest date at which anything relevant could have been created?
5. What would settle the case, and what would we need to know to advise on it?
6. Which assumptions in the current strategy have been tested rather than guessed?

People

1. Who are the likely key custodians?
2. Which individuals actually communicated about the issues?
3. Who appears in the communication analysis but not on our custodian list?
4. Which custodians have left, and what happened to their accounts and devices?
5. Which shared mailboxes, group accounts or assistants handled relevant traffic?
6. Which third parties hold copies or logs?

Data

1. What systems existed during the relevant period, and have they changed?
2. How much information exists, by source and custodian?
3. How much is duplicated, and under which deduplication scope?
4. Which file types dominate, and which are hard to review (spreadsheets, images, CAD, audio)?
5. What data has already been deleted, and what is subject to automatic deletion?
6. Which repositories have not yet been considered?
7. Are relevant communications occurring outside corporate email?
8. Are historical periods held only in backups or archives?

Technology

1. Which platforms allow in-place search and export, and with what limitations?
2. What are the processing settings and exception rates?
3. Do the proposed search terms return manageable, sampled, defensible results?
4. Is the population suitable for TAR or CAL, and how will recall be measured?
5. If generative AI is used, how are confidentiality, verification and audit trail secured?
6. Can the analysis be reproduced and explained to the opponent and the court?

Risk

1. Which communications contain potentially important evidence?
2. Where are the adverse documents, and who has been told?
3. Is there any indication of deletion, exfiltration or manipulation?
4. Which custodians used personal devices or accounts for work?
5. Are timestamps or authorship likely to be disputed?
6. What is the preservation risk register, and who owns each item?

Cost

1. What will collection cost, by source and method?
2. What will processing cost, given the expansion ratio?
3. What will hosting cost per month, and for how long?
4. What will review cost, on the measured population and an assumed rate?
5. Can the matter be staged so that later spend depends on earlier findings?
6. What does the CPR Part 3 costs budget say, and does it match the measured data?

Disclosure

1. Which Disclosure Model is realistic for each issue given what the data shows?
2. Which custodians, date ranges and sources will we propose in DRD section 2, and on what evidence?
3. Which sources are unavailable or inaccessible, and how will we describe them?
4. What will we tell the opponent about technology and sampling, and when?
5. Which collection activities are proportionate now, and which are deferred?
6. What evidence is missing, and where might alternative copies exist?

Privilege

1. Which lawyers, law firms and legal domains appear in the data?

2. Which document families are likely privileged, and are they isolated from wide review?
3. Is any assessment output itself privileged, and is it stored accordingly?
4. What is the protocol for inadvertent disclosure of privileged material?

Data protection

1. What personal data is being collected, and is each source necessary?
2. Is special-category data present (health, union membership, ethnicity), and how is it protected?
3. Which employees and third parties should be informed, and when?
4. Where is the data hosted, who has access, and for how long will it be retained?
5. Have data-minimisation choices (custodians, dates, sources) been recorded?
6. Is any transfer outside the UK involved, and on what basis?

§ 15 · OUTPUTS

ECA deliverables and sample templates

A professional ECA exercise produces documents, not impressions. The templates below are skeletons; each is populated with measured figures and the date on which they were measured.

EVIDENCE-SOURCE MAP
Source · Owner · Location · Custodians · Period covered · Access method · Export limits · Preservation status · Priority

CUSTODIAN MATRIX
Name · Role · Dates in role · Status (current/left) · Devices · Accounts · Aliases · Sources · Interviewed (Y/N) · Priority tier · Basis

DATA-VOLUME REPORT
Source · GB collected · Items · Documents after expansion · After dedupe · Exceptions · Date measured

FILE-TYPE ANALYSIS
Extension family · Count · Percentage · Reviewability note (e.g. spreadsheets, images requiring OCR, CAD)

DATE DISTRIBUTION
Month · Documents · Emails · Chats · Peak flag · Gap flag · Comment

COMMUNICATION MAP
From · To · Count · Channel · Internal/external · Domain · Rank

DOMAIN ANALYSIS
Domain · Messages · Custodians involved · Category (customer, competitor, personal, legal, newsletter) · Action

KEY-CONCEPT REPORT
Cluster label · Documents · Sample reviewed · Relevance verdict · Follow-up

KEYWORD-TEST REPORT
Term · Hits · Unique hits · Families · Sample size · Sample relevance rate · Keep/modify/drop · Reason

DUPLICATION REPORT
Scope (global/custodial) · Exact duplicates · Near-duplicate groups · Thread reduction · Suppressed items retained (Y)

TIMELINE
Date/time (UTC) · Event · Source document ID · Custodian · Confidence · Issue reference

HOT-DOCUMENT LIST
Doc ID · Date · Author · Summary · Issue · Helpful/adverse/neutral · Escalated to · Date

ADVERSE-DOCUMENT ESCALATION
Doc ID · Nature of concern · Reviewed by · Supervising lawyer decision · Client informed (date) · Privilege status

DATA-GAP REPORT
Expected evidence · Why expected · Primary source · Status · Alternative paths · Owner · Deadline

PRESERVATION-RISK REGISTER
Source · Risk (auto-delete, leaver, backup expiry) · Days until loss · Mitigation · Owner · Done (date)

COLLECTION RECOMMENDATION
Source/custodian · Collect now / defer / preserve only · Method (forensic image, in-place export, API) · Basis · Est. GB

DISCLOSURE-COST ESTIMATE
Stage · Volume driver · Unit assumption · Estimate · Range · Basis · Date

REVIEW STRATEGY
Population · Priority order · Method (linear, CAL, hybrid) · Team · QC sample · Recall target · Privilege workflow

TECHNICAL-RISK REGISTER
Risk (encryption, proprietary format, corrupt backup, export cap) · Impact · Likelihood · Mitigation · Owner

§ 16 · DASHBOARD CONCEPT

ECA dashboard concept

A mock dashboard showing the kind of intelligence a good ECA platform surfaces. Every value is fictional and illustrative; none is drawn from a real matter or from any benchmark.

ECA DASHBOARD · MATTER HS-2026-014 · HALDEN SYSTEMS (HYPOTHETICAL) — ILLUSTRATIVE DATA ONLY

18	31	24	3
CUSTODIANS	POTENTIAL SOURCES	SOURCES ASSESSED	PRESERVATION RISKS OPEN

412 GB	1.62m	0.94m	2021-03 to 2026-06
SOURCE VOLUME	PROCESSED DOCUMENTS	AFTER DEDUPLICATION (GLOBAL)	DATE RANGE · PEAK 2026-02 TO 2026-05

MOST ACTIVE CUSTODIANS	DOCS	EXTERNAL
Voss (sales director)	188,400	44%
Okafor (account manager)	121,900	39%
Reyes (customer success)	97,300	21%
Lindqvist (CEO)	63,100	12%

TOP EXTERNAL DOMAINS	MSGS	CATEGORY
northgate-cnc.example	2,310	customer
arcvale.example	1,940	competitor
gmail.example	1,120	personal
hartleylaw.example	410	legal

ISSUES	COUNT	STATUS
Hot documents	37	reviewed
Potential adverse documents	9	escalated
Potentially privileged (family)	2,860	isolated
Missing-source warnings	4	WhatsApp group; 2019 archive; Voss iPhone; CRM audit pre-2024

COST ESTIMATES (ILLUSTRATIVE)	VALUE	BASIS
Collection	£A	sources x method
Processing	£B	412 GB x rate
Review population	138,000	after threading and terms
Review hours	2,760	at 50 docs/hour (assumed)

Communication spike: 9x baseline external traffic from Voss between 14 and 27 May. Warning: Teams chat retention policy 90 days; hold applied 2 June.

Everything on this dashboard is invented to illustrate the categories of intelligence ECA can provide: matter overview, data, communications, issues and costs. "£A" and "£B" stand for estimates the provider would populate from measured volumes and quoted rates.

§ 17 · PRIVILEGE AND CONFIDENTIALITY

ECA and privilege

Early assessment identifies where privileged material concentrates before a wide review team sees it. Domain analysis lists the law firms, in-house legal addresses and legal department groups in the data; name lists identify

lawyers and paralegals; family analysis shows the email chains and attachments that travelled with legal advice; date analysis shows the period after litigation was contemplated, when litigation privilege may attach.

Why privilege searching is not infallible

Privilege is a legal characteristic of a communication, not a property of a domain. Lawyers send non-privileged emails (diary invitations, invoices, commercial negotiation as agent); non-lawyers forward privileged advice in ways that retain privilege, and sometimes in ways that lose it; in-house lawyers wear commercial hats; legal advice is summarised in board minutes; and litigation privilege turns on dominant purpose, which no search term measures. Automated privilege flagging, including generative AI flagging, is a triage tool that must be followed by lawyer review of every document to be withheld or produced.[12]

Safeguards before substantive material is exposed too widely

- Run domain, name and family analysis on the ECA dataset before any reviewer outside the core team sees content; isolate hits into a privilege workflow.
- Restrict access to the ECA platform by role; log who has viewed what.
- Treat ECA work product (hot-document lists, chronologies, assessments) as privileged: create it for the purpose of obtaining legal advice or for the litigation, label it, and store it outside the disclosure database.
- Agree a clawback or inadvertent-disclosure protocol with the opponent early; note CPR 31.20 on inadvertently disclosed privileged documents.[2]
- Where a provider or expert hosts the data, ensure the engagement terms address confidentiality, privilege and no secondary use of the data (including for model training).

LAWYER'S QUESTION

Does running analytics on the client's data risk waiving privilege?

TECHNICAL ANSWER Analytics do not disclose anything to anyone; they index and score. Waiver risk arises from how outputs are shared and stored, not from the computation.

PRACTICAL CONSEQUENCE Keep assessment outputs in a privileged workspace, control distribution, and think before pasting an analytics screenshot into an open letter to the opponent.

§ 18 · DATA PROTECTION

ECA and UK data protection

Almost every ECA dataset is personal data. The UK GDPR and the Data Protection Act 2018, as amended by the Data (Use and Access) Act 2025, govern its collection, hosting, analysis, transfer and retention. The framework does not prevent disclosure; it disciplines how disclosure is prepared.[15,16,17]

CONSIDERATION	POSITION (ENGLAND AND WALES, AS AT SEPTEMBER 2026)
Lawful basis	Processing for the purposes of actual or prospective legal proceedings is ordinarily justified under Article 6(1)(f) (legitimate interests) or, where a court order applies, Article 6(1)(c). The DUAA 2025 introduced "recognised legitimate interests" in a new Article 6(1)(ea) and Annex 1 for specified purposes; legal claims were already well supported and practitioners should check the current text of Annex 1 rather than assume it applies.[15,17]
Legal claims and legal proceedings	Article 9(2)(f) permits processing of special-category data where necessary for the establishment, exercise or defence of legal claims; DPA 2018 Schedule 1 para 33 and Schedule 2 para 5 provide the legal-claims condition and exemptions from certain provisions where disclosure is required by law or court order or is necessary for legal proceedings.[15,16]

Data minimisation and proportionality	Article 5(1)(c) requires data to be adequate, relevant and limited to what is necessary. Collecting whole devices and mailboxes "just in case" processes large quantities of irrelevant personal data of employees, families and third parties. Proportionate scoping recorded in the ECA workpapers is the practical evidence of compliance.[15,18]
Special-category data	Health, union membership, ethnicity, sexual orientation and similar data will appear in HR threads and personal messages. Identify it during ECA, apply Article 9 and the Schedule 1 conditions, restrict access and redact where it is irrelevant.[15,16,18]
Security	Article 32 requires appropriate technical and organisational measures. Assessment platforms, forensic images and exports must be encrypted, access-controlled and logged; chain-of-custody documentation serves both evidential and security purposes.
Retention	Preserved data is retained for the life of the matter; assessment copies that are not promoted to the disclosure set should be scheduled for deletion or return under the engagement terms.
Transfers	Hosting or review outside the UK requires an adequacy basis, the UK IDTA or Addendum, or another Chapter V mechanism. Confirm the provider's hosting location before data moves.[15]
Transparency	Employees will usually have been told in privacy notices that their data may be used for legal proceedings; check the notices, and consider whether specific information is needed when personal devices or personal accounts are collected.[18]

TWO THINGS ECA DOES NOT PERMIT

First, it does not permit a party to avoid disclosure obligations by citing data protection: where disclosure is ordered or required by the rules, the legal-proceedings provisions apply and the obligation is met with appropriate safeguards (redaction, confidentiality rings, protective orders). Second, it does not permit narrowing preservation on data-minimisation grounds: the duty to preserve documents that may be relevant is prior to, and independent of, any later collection scoping.[3,16]

§ 19 · COMMON MISTAKES

Fifteen Early Case Assessment mistakes that can become expensive later

1. Starting too late

PROBLEM ECA begins after collection, or after the DRD is drafted.

WHY IT MATTERS The decisions ECA should inform have already been taken; the DRD contains guesses that the data will contradict.

BETTER APPROACH Start ECA the week the dispute is identified, alongside preservation; treat the DRD as an output of ECA, not a trigger for it.

2. Confusing preservation with collection

PROBLEM "We have not collected X" is treated as "X is not preserved", or vice versa.

WHY IT MATTERS Under-preservation breaches PD 57AD para 3.1; over-collection wastes money and processes unnecessary personal data.[3]

BETTER APPROACH Preserve broadly and immediately; collect in evidenced stages; record both separately.

3. Collecting before understanding systems

PROBLEM Devices are imaged and tenants exported before anyone knows what the systems hold or how they store it.

WHY IT MATTERS Teams chat, SharePoint versions and CRM audit trails each require specific collection methods; a generic export may miss or mangle them.

BETTER APPROACH Interview IT and custodians, map sources, and choose method per source.

4. Choosing custodians solely from an organisation chart

PROBLEM Seniority is used as a proxy for relevance.

WHY IT MATTERS The people who did the work and sent the messages are often two levels down; senior mailboxes are large and thin.

BETTER APPROACH Run communication analysis on an early dataset and revise the custodian tiering from the data.

5. Ignoring former employees

PROBLEM Leavers are omitted because "they are not here to ask".

WHY IT MATTERS Leaver processes delete accounts on a schedule; the most relevant custodian in a data-theft case has, by definition, left.

BETTER APPROACH Identify leavers in the relevant period on day one; suspend deletion; locate inactive mailboxes and reissued devices.

6. Relying only on interviews

PROBLEM Custodian recollection is treated as a complete inventory.

WHY IT MATTERS People forget channels, misremember dates and, occasionally, omit what embarrasses them.

BETTER APPROACH Use interviews to generate hypotheses and metadata to test them.

7. Assuming email contains everything

PROBLEM Scope is defined as mailboxes plus date range.

WHY IT MATTERS Substantive discussion has migrated to chat, CRM notes, shared documents and messaging apps; email is often the least candid channel.

BETTER APPROACH Measure channel split in the ECA dataset and follow the evidence.

8. Ignoring Teams, Slack and WhatsApp

PROBLEM Chat is seen as informal and therefore unimportant, or as too hard to collect.

WHY IT MATTERS Chat retention is short, export is fiddly, and the DRD expressly lists instant messaging and collaboration systems as sources to address.[4]

BETTER APPROACH Preserve chat immediately; collect it with tools that keep conversation structure; treat handsets as sources.

9. Using untested keywords

PROBLEM Terms are agreed with the opponent before being run against the data.

WHY IT MATTERS Over-inclusive terms create the review population; under-inclusive terms miss documents and invite a further-search order (Digicel).[7]

BETTER APPROACH Test every term for hits, families and sampled relevance, and negotiate from the results.

10. Ignoring metadata

PROBLEM Documents are assessed by content alone.

WHY IT MATTERS Authorship, edit history, transmission and time-zone data are often the evidence; careless handling destroys them.

BETTER APPROACH Collect forensically, record time-zone settings, and read the metadata reports.

11. Treating deduplication as merely deletion

PROBLEM Duplicates are discarded and the fact of custody is lost.

WHY IT MATTERS Who held a document, and when, may be the issue; suppressed copies must remain traceable.

BETTER APPROACH Record the deduplication scope, retain "all custodians" fields, and keep suppressed items in the dataset.

12. Failing to validate samples

PROBLEM A model's prediction is reported as a finding without human sampling, tolerances or rounds.

WHY IT MATTERS This is precisely the failing in Triumph Controls at [29] and [36].[6]

BETTER APPROACH State sample size, confidence level and margin in advance; have humans review the sample; report overturns.

13. Overlooking privileged material

PROBLEM Wide review teams see legal advice before privilege is screened.

WHY IT MATTERS Inadvertent production and confidentiality breaches follow.

BETTER APPROACH Domain and name screening in ECA, privilege workflow, clawback protocol (section 17).

14. Ignoring cloud retention and deletion

PROBLEM Nobody checks the tenant's retention labels, chat policies or backup rotation.

WHY IT MATTERS Data is lost silently while lawyers argue about scope; the loss then has to be explained.

BETTER APPROACH Make retention review the first ECA task; document holds applied and by whom.

15. Failing to document decisions

PROBLEM Scope decisions live in emails and memories.

WHY IT MATTERS Months later the opponent asks why a custodian was excluded; the answer must be evidenced (Triumph Controls at [29]).[6]

BETTER APPROACH Maintain a decision log: decision, date, evidence relied on, decision-maker, review date.

§ 2 0 · WHEN TO INVOLVE A DIGITAL FORENSIC EXPERT

When digital forensics should enter the ECA process

Ordinary eDiscovery collection copies what exists, in a defensible way, for review. Forensic investigation asks what happened: what existed and no longer does, who did what on which device and when, and whether artefacts are what they appear to be. The two disciplines overlap at collection and diverge at analysis; the point of ECA is to notice, early, which one the facts require.

CIRCUMSTANCE	WHY FORENSIC EXPERTISE IS NEEDED EARLY
Deleted information or suspected data destruction	Recovery from unallocated space, recycle bins, shadow copies and cloud retention is time-limited; forensic imaging preserves the possibility, and the audit trail proves the act.
USB activity, data exfiltration, personal cloud indicators	Registry, setupapi logs, LNK files, browser and sync-client artefacts establish what was copied, when and where; eDiscovery exports do not capture them.
Personal devices and mobile applications	WhatsApp, Signal and iMessage require handset-level extraction; consent, proportionality and scoping must be handled carefully.
Damaged or encrypted computers	Recovery and lawful decryption need specialist tools; a RAM capture at seizure may hold the key (see section 4.13 of our guide Digital Forensic Evidence).
Disputed timestamps or manipulation allegations	File-system, application and cloud metadata are compared and clock drift assessed; authenticity of images, audio and documents is tested.
User attribution on shared or multi-user systems	Logon events, profile activity and application traces establish who was at the keyboard.
Cloud deletion, departing employees and insider threat	Unified audit logs, sign-in logs and inactive-mailbox mechanics must be secured before retention windows close.
Missing documents and preservation of volatile evidence	The alternative-evidence-path exercise in section 9 is forensic work when the primary source has gone.

Where the matter may reach a criminal court, or where an expert report may be relied on in criminal proceedings, the Forensic Science Regulator's statutory Code applies and compliance declarations are required; in civil proceedings the expert's overriding duty to the court under CPR Part 35 applies.[22,23] Instruct with a clear question, a chain-of-custody expectation and a prohibition on anyone else powering on or examining the source before the examiner does.

§ 2 1 · IMPLEMENTATION PLAN

Practical implementation plan

WINDOW	LEGAL TEAM	CLIENT IT	EDISCOVERY TEAM	DIGITAL FORENSIC TEAM

First 24 hours	Issue hold notice; identify obvious custodians and leavers; open decision log; instruct providers; write the ten questions.	Suspend deletion routines, leaver purges and chat retention; list systems and tenants; freeze reissue of devices.	Open matter; agree platform, hosting location and security; request retention and tenant configuration.	Advise on volatile sources; take custody of any at-risk device; capture logs with short rotation.
First 72 hours	Interview key custodians; draft evidence-source map; assess adverse-evidence risk; brief client on preservation duties.	Confirm holds in place; export retention and audit configurations; identify backups and archives.	Stage one in-place searches and metadata pulls on priority custodians; begin volume report.	Image priority devices where attribution or deletion is in play; secure handsets by consent or order.
First week	Strategy meeting on ECA findings; revise issues and custodian tiers; consider settlement posture; instruct on gaps.	Provide access for targeted collection; answer system-history questions; locate legacy data.	Build ECA dataset; run analytics; test terms; produce date, domain and communication reports; draft cost ranges.	Report on artefacts (USB, cloud sync, deletion); identify alternative evidence paths for missing items.
Before broad collection	Approve collection recommendation and defer list; record rationale; consider data-protection scoping.	Provision access; confirm export capabilities and limits per source.	Finalise collection methods; sample deferred sources; quantify expansion and duplication.	Forensic collection for sources needing chain of custody; witness statements on method.
Before the DRD	Draft DRD section 2 from measured data: sources, custodians, dates, unavailable sources, technology, costs; align with costs budget.	Confirm system inventory statements are accurate; sign off accessibility notes.	Provide figures and technology proposal; prepare explanation of deduplication and threading.	Provide accessibility and integrity notes for damaged, encrypted or legacy sources.
Before agreeing search methodology	Share tested-term results and sampling with the opponent; propose TAR/CAL where proportionate; agree protocol, tolerances and reporting.	Confirm no configuration changes since preservation.	Run agreed terms; report hit and family counts; set up validation sampling.	Confirm forensic sources are processed with metadata intact; stand by for authenticity disputes.

§ 2 2 · QUESTION BANKS

Question banks

Questions to ask the client

1. Who worked on this matter day to day, and who did they talk to? Who has left, and when?
2. Which systems were in use during the relevant period, and have any been migrated, retired or replaced?
3. What retention, auto-delete and leaver policies apply to mailboxes, chat, files and backups? Have they been suspended?
4. Do people use WhatsApp, Signal, personal email or personal devices for work? Who, and for what?
5. Which business applications (CRM, ERP, finance, ticketing) hold records of the events?
6. What has already been deleted, wiped, reissued or reformatted since the dispute arose?
7. Are there backups or archives that reach back to the relevant period? Who can restore them?
8. Who are your lawyers and advisers, internal and external, and which addresses did they use?
9. Is there any document you are worried about? Who else has seen it?
10. What have you already told the other side, and what have you promised to do?

Questions to ask the opponent

1. Which custodians, date ranges and sources do you propose, and on what evidence?
2. What preservation steps have you taken, when, and for which systems, including chat and mobile?
3. Which sources are unavailable or inaccessible, and why?
4. Have you tested your proposed keywords? Please provide hit and family counts and sampling results.
5. Do you intend to use deduplication, threading, analytics, TAR or generative AI? Please describe the method, oversight and validation.
6. Will you share your processing settings (deduplication scope, time zone, exception handling)?
7. Which former employees' data have you preserved, and how?
8. What is your estimate of documents by source, and of disclosure cost?
9. Are any relevant communications held on personal devices or accounts of your witnesses?
10. Will you agree a clawback protocol and a production format now?

Questions to ask a forensic or eDisclosure provider

1. How will you preserve and collect each source, and how is chain of custody documented?
2. What can you assess in place before full collection, and what does that cost?
3. What is your deduplication scope by default, and how are suppressed items retained and reported?
4. How do you handle time zones, exceptions, encrypted items and non-text files?
5. What analytics are available for ECA, and what reports will we receive (volume, date, domain, communication, concept, keyword-test)?
6. If TAR or CAL is proposed, how is recall measured and reported, and who trains the model?
7. If generative AI is offered, where is data processed, is it used for training, and how are outputs verified and logged?
8. Where is data hosted, who has access, and what security certifications apply?
9. Can you give evidence about your method, and have you done so before?
10. What are your unit rates, and what volume assumptions drive the estimate?

§ 2 3 · SUGGESTED WORDING

Model instruction wording

Adapt to the matter; these are starting points, not precedents.

A · INSTRUCTION TO AN EDISCLOSURE PROVIDER FOR AN ECA EXERCISE

"We instruct you to carry out an early data assessment in relation to [matter]. Please (1) preserve, in a forensically sound manner and with full chain-of-custody documentation, the sources listed in Schedule 1; (2) collect, using the methods specified in Schedule 2, the priority custodians and sources in Schedule 3 for the period [dates], without prejudice to later collection of the remaining preserved sources; (3) process the collected data with global deduplication, retaining all-custodian fields and suppressed items, using UTC with [time zone] display; (4) provide within [n] days a data-volume report, date-distribution report, file-type analysis, domain analysis, communication map, keyword-test report for the terms in Schedule 4 with sampled relevance rates, and a data-gap report; (5) escalate immediately to [supervising lawyer] any document that appears adverse to our client's case; and (6) treat all work product as privileged and confidential, hosted in [location], with no secondary use of the data."

B · PRESERVATION INSTRUCTION TO THE CLIENT

"With immediate effect, please suspend all automatic deletion, retention expiry, archiving and leaver-account purge processes applying to the mailboxes, Teams and other chat data, OneDrive and SharePoint content, devices and backups of the individuals and systems listed in the attached schedule, and confirm in writing when this has been done and by whom. Do not power on, examine, reissue, reimage or dispose of any listed device. This instruction covers former employees' data and any personal devices or accounts used for work. Preservation is broader than the material we will later collect; please do not narrow it without our written agreement."

C · LETTER TO THE OPPONENT ON ECA-INFORMED SCOPE

"Our client has assessed its data sources for the purpose of completing section 2 of the DRD. Communication analysis across [n] custodians for the period [dates] shows that [x] per cent of documents responsive to the draft Issues for Disclosure were exchanged between [custodians]; the date distribution is enclosed. We propose that searches be limited under PD 57AD para 9.6 to the custodians, date ranges and repositories set out in the enclosed schedule, and that deduplication and email threading be applied on the settings described. Our client has preserved, and will continue to preserve, the remaining sources listed. If you contend that any further custodian or source should be searched, please identify it and the basis for that contention so that the point can be considered before the CMC."

D · INSTRUCTION TO A DIGITAL FORENSIC EXAMINER WITHIN ECA

"Please take custody of and forensically image the devices listed in Schedule 1 and secure the logs in Schedule 2, and report on (a) any indication of deletion, wiping or reinstallation after [date]; (b) USB and removable-media connections and files accessed from them; (c) installation or use of personal cloud-storage or file-transfer applications and any synchronisation activity; (d) the authenticity and creation history of the documents in Schedule 3; and (e) any alternative sources from which material apparently deleted may be recovered. Your report should comply with CPR Part 35 and, if the matter may be referred to the police, with the Forensic Science Regulator's Code. Please do not alter or examine any source until imaging is complete and hash-verified."

§ 24 · RED FLAGS

Red flags

Signals that the ECA exercise, or the disclosure that depends on it, is heading towards a dispute or a further-search order.

- The custodian list has not changed since the first client call.
- Keywords have been agreed with the opponent but never run against the data.
- Nobody can state the deduplication scope or the time zone applied.
- A single-digit percentage relevance figure is quoted for an unreviewed population with no sample size or method.
- The DRD says "manual review" or "TAR" and the team intends to do something else.
- Emails refer to a WhatsApp group, a Signal chat or "my personal Gmail" and no handset has been considered.
- The most relevant employee has left and nobody has checked the leaver process.
- The date distribution has a hole in the middle of the critical period and nobody has asked why.
- A witness produces a document during proofing that is not in the disclosure set (Triumph Controls at [35]).[6]
- Analytics screenshots or hot-document lists are circulating outside the privileged workspace.
- The provider cannot say where the data is hosted or whether it is used to train models.
- Preservation has been "scoped down" to match the collection plan.
- Tenant retention policies were reviewed after collection, not before.
- The cost budget was prepared from source gigabytes rather than from a measured document population.
- Decisions about excluded sources exist only in email chains.

§ 25 · GLOSSARY

Glossary

Active learning / CAL

A TAR workflow in which the relevance model retrains continuously as reviewers code documents, presenting the likeliest-relevant documents next.

Adverse document
A document that is adverse to the disclosing party's case; under PD 57AD known adverse documents must be disclosed regardless of Model.

Chain of custody
The documented record of who held evidence, when, and what was done to it, from source to court.

Clustering
Statistical grouping of documents by textual similarity.

Concept search
Retrieval of documents by meaning or theme rather than by literal term.

Custodial deduplication
Removal of duplicates within each custodian's data only.

Custodian
A person (or account) who holds or controls potentially relevant documents.

Deduplication
Identification and suppression of exact duplicates by hash value.

Disclosure Model
One of Models A to E under PD 57AD para 8, from known adverse documents only (A) to wide search-based disclosure (E).

DRD
Disclosure Review Document under PD 57AD para 10.

ECA
Early Case Assessment: early analysis of the dispute informed by data.

EDA
Early Data Assessment: early analysis of the ESI itself.

EDQ
Electronic Documents Questionnaire under PD 31B.

EDRM
Electronic Discovery Reference Model, an industry framework; not English procedural law.

Email threading
Reconstruction of conversations and identification of inclusive messages.

ESI
Electronically stored information.

Expansion ratio
The multiplier from collected files to processed documents after containers and attachments are extracted.

Family
A parent document and its attachments, reviewed and produced together.

Forensic image
A bit-for-bit copy of a storage device, hash-verified against the original.

Generative AI
Large language models that produce text such as summaries, classifications and answers.

Hash value
A cryptographic fingerprint of a file; identical files produce identical hashes.

Hot document
A document of unusual strategic importance, helpful or adverse.

Inclusive message
The message in a thread that contains all earlier content.

Issues for Disclosure
The key issues in dispute to be determined by reference to contemporaneous documents (PD 57AD para 7).

Legal hold
An instruction to preserve documents and suspend deletion.

Metadata
Data about data: authorship, dates, recipients, file properties.

Near-duplicate
A document substantially similar but not identical to another.

Precision
The proportion of documents retrieved that are relevant.

Predictive coding
Earlier term for TAR; supervised relevance classification.

Proportionality
The overriding objective's requirement to deal with cases at proportionate cost (CPR 1.1).

Recall
The proportion of relevant documents actually retrieved.

Sampling
Review of a random or stratified subset to estimate properties of the whole.

Special-category data
Personal data listed in UK GDPR Article 9(1), such as health or ethnicity.

TAR
Technology-assisted review.

Unified audit log
Microsoft 365's consolidated record of user and admin operations.

§ 2 6 · CONCLUSION

Early knowledge changes the economics and strategy of disclosure

Well-designed ECA moves a legal team from assumptions about a dataset to measurable information about people, evidence sources, data volumes, relationships, important events, relevant documents, adverse evidence, missing evidence, costs and technical risks. It does not guarantee lower costs, and it does not guarantee better outcomes. What it provides is evidence-based intelligence at the one point in a case when legal and disclosure strategies can still be changed at modest cost.

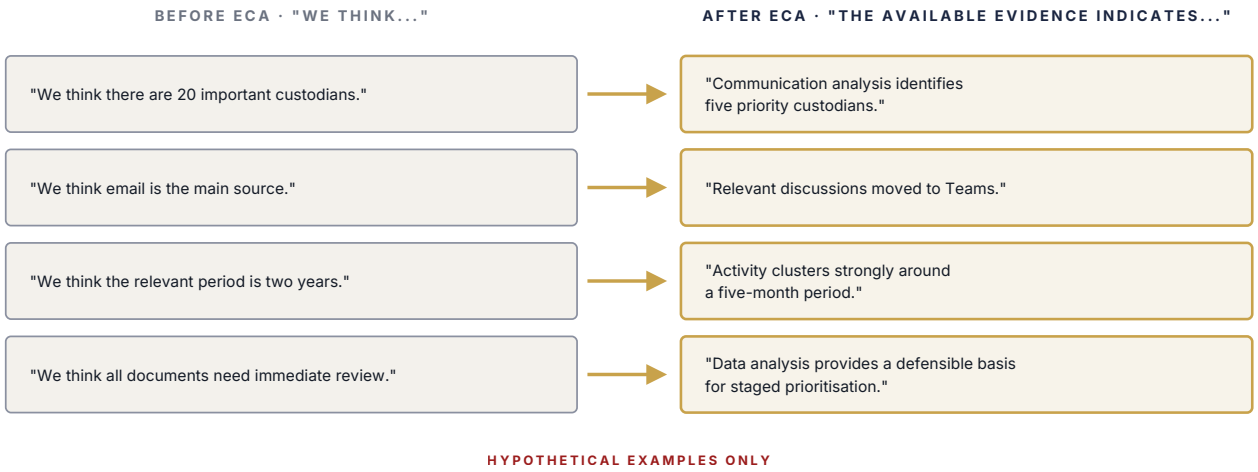


Figure 12 · What ECA gives the lawyer. Four assumptions and the evidence-based statements that replace them. Alt text: two columns of four boxes; grey "we think" statements on the left, gold-bordered evidence statements on the right, joined by arrows.

When specialist eDiscovery and digital forensic support may help

Many legal teams run the first stages of ECA in-house. External specialists tend to add value where the exercise needs tools, capacity or independence the team does not have: **evidence mapping** across unfamiliar systems; **preservation** of volatile and cloud sources; **forensic collection** where chain of custody, attribution or deletion may be in issue; **cloud collection** from Microsoft 365, Google Workspace and SaaS platforms with their export limits; **processing** with documented, defensible settings; **analytics** and the reports described in section 15; **ECA** itself as a scoped, time-boxed engagement; **review technology** including TAR/CAL with validated recall reporting; **forensic investigation** where the facts require it; **production** to agreed protocols; and **expert evidence** on method under CPR Part 35 or CrimPR Part 19. Computer Forensics Lab and its eDiscovery arm, e-discovery.uk, provide each of these from a single London laboratory; details are at cflab.uk and on the final page. The best measure of any provider is whether its ECA report changes what the legal team does next.

§ 27 · REFERENCES AND BIBLIOGRAPHY

References and bibliography

All links accessed 1 September 2026. Neutral citations and paragraph references are to the official transcripts on the National Archives Find Case Law service.

1. Ministry of Justice, Civil Procedure Rules Part 1: Overriding Objective and Part 3: The Court's Case and Costs Management Powers (current edition). justice.gov.uk/courts/procedure-rules/civil/rules/part01; .../part03.
2. Ministry of Justice, Civil Procedure Rules Part 31: Disclosure and Inspection of Documents and Practice Direction 31B: Disclosure of Electronic Documents (current edition). justice.gov.uk/courts/procedure-rules/civil/rules/part31; .../part31/pd_part31b.
3. Ministry of Justice, Practice Direction 57AD: Disclosure in the Business and Property Courts (in force 1 October 2022, as amended; current text checked 1 September 2026). justice.gov.uk/.../practice-direction-57ad-disclosure-in-the-business-and-property-courts.
4. Ministry of Justice, Disclosure Review Document (Appendix 2 to PD 57AD) and Explanatory Notes for the DRD (current versions). justice.gov.uk/__data/assets/pdf_file/0009/177471/disclosure-review-document.pdf.
5. Pyrrho Investments Ltd & Anor v MWB Property Ltd & Ors [2016] EWHC 256 (Ch), High Court (Chancery Division), Master Matthews, 16 February 2016, at [2] to [5], [15], [17] to [24], [33] to [34]. caselaw.nationalarchives.gov.uk/ewhc/ch/2016/256.
6. Triumph Controls UK Ltd & Anor v Primus International Holding Co & Ors [2018] EWHC 176 (TCC), High Court (Technology and Construction Court), Coulson J, 7 February 2018, at [1], [3], [6] to [11], [15] to [18], [19] to [24], [25] to [42]. caselaw.nationalarchives.gov.uk/ewhc/tcc/2018/176.
7. Digicel (St Lucia) Ltd & Ors v Cable & Wireless Plc & Ors [2008] EWHC 2522 (Ch), Morgan J, 23 October 2008, at [81], [94] to [96] (as quoted in Triumph Controls at [13] to [14]). caselaw.nationalarchives.gov.uk/ewhc/ch/2008/2522.
8. Nichia Corporation v Argos Ltd [2007] EWCA Civ 741, Court of Appeal, at [50] to [52] (as quoted in Triumph Controls at [12]). caselaw.nationalarchives.gov.uk/ewca/civ/2007/741.
9. Goodale v Ministry of Justice [2009] EWHC B41 (QB), Senior Master Whitaker, at [1] to [4], [26] to [27] (as quoted in Pyrrho at [13] to [14]).
10. EDRM, EDRM Identification Standards (Early Case Assessment and Early Data Assessment as the two components of identification), edrm.net, published under CC BY 4.0. edrm.net/.../edrm-identification-standards/.
11. EDRM, Early Case Assessment: Evolving from Tactical to Practical (white paper; statistical, date, textual and relationship analytics in ECA). edrm.net/papers/early-case-assessment-evolving-from-tactical-to-practical/. Industry material; used for taxonomy only.
12. The Law Society of England and Wales, Generative AI in legal disclosure: a practical guide (November 2025). lawsociety.org.uk/topics/civil-litigation/generative-ai-in-legal-disclosure-guide. See also Generative AI: the essentials (May 2025). lawsociety.org.uk/topics/ai-and-lawtech/generative-ai-the-essentials.
13. International Legal Technology Association (ILTA), Active Learning Best Practice Guide (2024) and Generative AI Best Practice Guide (30 September 2025), for litigators in England and Wales. Professional guidance; available via ilta.net and reported at legaltechnology.com (25 April 2025).
14. Ayinde v London Borough of Haringey; Al-Haroun v Qatar National Bank [2025] EWHC 1383 (Admin), Divisional Court, June 2025 (professional accountability for AI-generated material). caselaw.nationalarchives.gov.uk/ewhc/admin/2025/1383.
15. UK General Data Protection Regulation (Regulation (EU) 2016/679 as retained and amended), Articles 5(1)(c), 6, 9(2)(f), 32 and Chapter V. legislation.gov.uk/eur/2016/679/contents.
16. Data Protection Act 2018, Schedule 1 para 33 (legal claims) and Schedule 2 para 5 (disclosure required by law or in connection with legal proceedings). legislation.gov.uk/ukpga/2018/12/contents.
17. Data (Use and Access) Act 2025 (amendments to UK GDPR and DPA 2018, including recognised legitimate interests; provisions commenced in stages through 2025 and 2026). legislation.gov.uk/ukpga/2025/18/contents.
18. Information Commissioner's Office, Guide to the UK GDPR: Principles (data minimisation); Special category data; Legitimate interests; and Data protection and journalism / legal claims materials (current versions). ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/.
19. TeCSA, TECBAR and SCL, eDisclosure Protocol (Technology and Construction Court; paras 4.2, 5.2 and 5.4 as cited in Triumph Controls at [16]). tecsa.org.uk.
20. Grossman, M. R. and Cormack, G. V., "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review" (2011) 17 Richmond Journal of Law and Technology 11. Independent research; cited through Irish Bank Resolution Corporation v Quinn [2015] IEHC 175 at [66], quoted in Pyrrho at [31]. jolt.richmond.edu/jolt-archive/v17i3/article11.pdf.
21. Irish Bank Resolution Corporation Ltd v Quinn [2015] IEHC 175, High Court of Ireland, Fullam J, at [66] to [69] (as quoted in Pyrrho at [31]). Persuasive only.

22. Forensic Science Regulator, Code of Practice, version 2 (statutory; in force 2 October 2025) and associated guidance FSR-GUI-0001 Issue 3 (March 2026). gov.uk/government/publications/forensic-science-regulator-code-of-practice.
23. Ministry of Justice, Civil Procedure Rules Part 35: Experts and Assessors; Criminal Procedure Rules Part 19: Expert Evidence; Attorney General's Office, Guidelines on Disclosure (2024); Criminal Procedure and Investigations Act 1996. justice.gov.uk/.../part35; gov.uk/.../attorney-generals-guidelines-on-disclosure.
24. Brown v BCA Trading Ltd & Ors [2016] EWHC 1464 (Ch), Registrar Jones, 17 May 2016 (predictive coding ordered where parties disagreed). caselaw.nationalarchives.gov.uk/ewhc/ch/2016/1464.
25. Computer Forensics Lab, Digital Forensic Evidence: What Every Defence Lawyer Needs to Know and the CFL Electronic Evidence Series (Guides 1 to 4: Electronic Disclosure Explained; eDisclosure or Digital Forensics?; The EDRM for UK Litigation Teams; Finding the Evidence). cflab.uk; e-discovery.uk.

§ 28 · DISCLAIMER

Disclaimer and jurisdictional notes

This guide is published by Computer Forensics Lab Ltd for general educational purposes. It provides general information about early case assessment, electronic disclosure and digital forensics in England and Wales as at 1 September 2026. It is not legal advice, does not create a solicitor-client or expert-client relationship, and should not be relied on as a substitute for advice from a qualified lawyer on the facts of a particular matter. Procedural rules, practice directions, guidance and case law change; readers should verify the current position before acting. The hypothetical matter, the parties, the individuals, the dashboard values and the cost figures in this guide are fictional and illustrative. Any resemblance to real persons, companies or matters is coincidental. Nothing in this guide should be read as a description of any actual client engagement.

FORUM	PRINCIPAL DIFFERENCES FROM THE ENGLAND AND WALES CIVIL POSITION DESCRIBED IN THIS GUIDE
Business and Property Courts	PD 57AD applies (subject to para 1.4 exclusions): Issues for Disclosure, Models A to E, DRD, Disclosure Guidance.
Other England and Wales civil courts	CPR Part 31 and PD 31B: standard disclosure, reasonable search, EDQ in multi-track cases.
Scotland	Court of Session and Sheriff Court: recovery by specification of documents and commission and diligence; no DRD; proportionality applied through the court's discretion.
Northern Ireland	Rules of the Court of Judicature (NI) 1980, Order 24; broader relevance-based discovery; no PD 57AD equivalent.
Arbitration	Governed by the arbitration agreement, institutional rules and tribunal orders; IBA Rules on the Taking of Evidence commonly adopted; document production is typically narrower and request-based.
Tribunals	Employment Tribunals apply their own procedure rules and the overriding objective; disclosure is ordered case by case; no search-based Models.
Criminal proceedings	CPIA 1996, the Code of Practice and the Attorney General's Guidelines on Disclosure (2024); reasonable lines of enquiry; digital material strategies; expert evidence under CrimPR Part 19 and the Forensic Science Regulator's Code.

§ APPENDIX · DETACHABLE CHECKLIST

Early Case Assessment Checklist for Litigation Lawyers

Tick each item when it has been done and recorded. "Recorded" means written in the decision log with date, evidence and decision-maker.

Issues

- ☐ Pleaded or anticipated issues listed and mapped to documentary questions
- ☐ Draft Issues for Disclosure prepared and reviewed against ECA findings
- ☐ The ten things the team most needs to know written down

Custodians

- ☐ Obvious custodians listed from interviews and organisation chart
- ☐ Custodian tiers revised from communication analysis
- ☐ Leavers, shared accounts and assistants identified
- ☐ Custodian matrix completed and dated

Systems

- ☐ Evidence-source map covering all eight regions
- ☐ System history (migrations, retirements) confirmed with IT
- ☐ Export capabilities and limits recorded per source

Preservation

- ☐ Hold notice issued and acknowledged
- ☐ Auto-delete, retention and leaver processes suspended and confirmed in writing
- ☐ Devices frozen; at-risk devices in custody
- ☐ Preservation-risk register open with owners and dates

Data volumes

- ☐ Source volume measured by custodian and source
- ☐ Expansion ratio and exception rate recorded
- ☐ Deduplication scope chosen and rate recorded

Date ranges

- ☐ Date distribution produced per source
- ☐ Peaks and gaps explained
- ☐ Proposed ranges evidenced and recorded

Communications

- ☐ Communication map and domain analysis produced
- ☐ Channel split (email, chat, mobile) measured
- ☐ External, personal and competitor domains categorised

Search terms

- ☐ Every term tested for hits, families and sampled relevance
- ☐ Variants, abbreviations, aliases and code words considered
- ☐ Non-hit population sampled

Analytics

- ☐ Threading, clustering and concept reports reviewed
- ☐ TAR/CAL suitability assessed; recall measurement planned
- ☐ Generative AI use, if any, controlled and logged

Adverse documents

- ☐ Escalation route defined and used
- ☐ Supervising lawyer and client informed
- ☐ Model A implications considered

Privilege

- ☐ Legal domains and names screened before wide review
- ☐ Privileged families isolated
- ☐ ECA work product held in privileged workspace
- ☐ Clawback protocol proposed

Data protection

- ☐ Lawful basis and legal-claims provisions identified
- ☐ Special-category data identified and protected
- ☐ Minimisation decisions recorded; hosting location confirmed

Missing evidence

- ☐ Data-gap report produced
- ☐ Each gap assigned an owner and deadline

Alternative evidence sources

- ☐ Alternative paths identified for each gap (section 9)
- ☐ Counterpart and third-party sources approached where appropriate

Cost

- ☐ Formula-based estimates from measured volumes
- ☐ Ranges and assumptions stated
- ☐ Aligned with CPR Part 3 costs budget

Collection strategy

- ☐ Collect-now, defer and preserve-only lists recorded with rationale
- ☐ Method chosen per source; forensic sources identified

Review strategy

- ☐ Priority order and method chosen
- ☐ QC sampling and recall targets set
- ☐ Privilege workflow integrated

Disclosure strategy

- ☐ Models proposed per issue on evidence
- ☐ DRD section 2 drafted from measured data
- ☐ Technology and sampling proposals shared with opponent

Documentation

- ☐ Decision log maintained and current
- ☐ Processing settings and holds recorded
- ☐ Provider reports filed with dates

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