



C O M P U T E R F O R E N S I C S L A B

London digital forensics & e-discovery · Established 2007

Native Documents, PDFs or Images: Which Production Format?

What Each Format Preserves, What Each Destroys, and How to Choose in the Protocol

Every production makes a format decision, and every format decision is an evidence decision. Native files carry their formulas, tracked changes, speaker notes and embedded metadata: everything guide 87 taught you to read: and resist pagination, redaction and citation. Imaged formats fix the page for the bundle and the redaction tool: and strip the machine testimony wholesale, flatten the spreadsheet into unreadable acreage, and quietly delete the hidden columns where the case was living. The right answer is per-document-type, agreed in the protocol before the arguments need it, with challenge routes preserved for the documents that turn out to matter. This guide maps the trade-offs for UK lawyers: what each format preserves and destroys, the document classes that force the choice, PD 57AD's default and how to draft around it, and the disputes that follow when the choice was made badly.

🕒 Estimated reading time: 24 min read

§ ABOUT THE AUTHOR

PREPARED BY COMPUTER FORENSICS LAB E-DISCOVERY TEAM

Computer Forensics Lab is a London-based digital forensics and electronic disclosure practice established in 2007. Its eDiscovery arm, **e-discovery.uk**, produces and receives disclosure in every format this guide discusses, drafts and negotiates format protocols, and handles the disputes: flattened spreadsheets, stripped metadata, unusable natives: that follow when format was chosen carelessly. The forensic side supplies the provenance and examination work that format questions escalate into.

ESTABLISHED 2007 · LONDON

ISO 17025-ALIGNED PROCEDURES

PRODUCTION & FORMAT ENGINEERING

PD 57AD PROTOCOL PRACTICE

CPR PART 35 EXPERT REPORTS

FULL CHAIN-OF-CUSTODY DOCUMENTATION

§ CONTENTS

In this guide

01	Executive summary
02	The problem in plain English: the format is an evidence decision
03	The three families: native, searchable PDF and image-plus-load-file
04	Document classes that force the choice: spreadsheets, presentations, chat and media
05	The rules layer: PD 57AD, PD 31B and what "native" means in an order
06	Redaction, confidentiality and the imaging exception
07	Receiving productions: acceptance checks and challenge routes
08	Source architecture: where else the evidence lives
09	Worked examples
10	Common mistakes and technical limitations
11	Questions to ask · Suggested wording
12	Checklist and red flags · When to involve a digital forensic expert
13	Frequently asked questions
14	Glossary · References · Disclaimer · How a specialist laboratory can assist

Executive summary

THE HEADLINE POINT: PRODUCE NATIVE BY DEFAULT, IMAGE ONLY FOR CAUSE, AND PRESERVE THE ROUTE BACK TO THE NATIVE FOR EVERY DOCUMENT THAT MATTERS

The three format families trade evidence against convenience. **Native production** (the file as it lives: .xlsx, .docx, .msg) preserves everything: formulas, tracked changes, hidden content, embedded metadata: and is the modern default under PD 57AD, which expects documents in the form in which they are ordinarily maintained. **Searchable PDF** fixes pagination and travels well, but is a conversion: embedded metadata, dynamic content and hidden layers are stripped or frozen at render time, per guide 87 §7's warning. **Image-plus-load-file** (TIFF/PDF images with extracted text and agreed metadata fields) is the redaction and bundle workhorse: the page fixed, the fields preserved alongside: at the cost of the native's interior. The working rule: **native by default; imaging for cause** (redaction, sealing, genuine display need), with the cause stated; spreadsheets, presentations and databases produced native *always* absent agreement, because their imaged forms are evidence destruction wearing a page number; and the protocol preserving, for any produced document, the right to the native original and its provenance when authenticity or interior content becomes an issue: guide 87 §11's clause doing exactly this. Format disputes are cheapest at the DRD table and dearest in the witness box: the choice is drafted, not defaulted.

- **Native is the default, not the exception:** the form ordinarily maintained, interiors intact: PD 57AD's expectation and this guide's rule.
- **Spreadsheets never image well:** formulas, hidden sheets and filters are the evidence: flattened acreage is destruction, not production.
- **Imaging is for stated cause:** redaction and confidentiality justify it per document, with the conversion flagged.
- **Load files carry the fields:** where images are used, the agreed metadata travels beside them: guide 87's set.
- **The route back stays open:** native originals and provenance obtainable on request for contested documents: drafted now, needed later.

The problem in plain English: the format is an evidence decision

A document is more than its printout. The spreadsheet's cells contain formulas that show how the number was made, hidden columns where the earlier figures still sit, and filters that decide what a reader sees; the Word file carries its tracked deletions and comments; the presentation its speaker notes; every file its embedded biography per guide 87. Print any of them: to paper, to PDF, to TIFF: and you keep the surface and discard the interior. Sometimes that is exactly right (the redacted page must be fixed); mostly it is a quiet subtraction nobody authorised.

Format therefore belongs with scope and search among the protocol's real decisions. Made deliberately, it is uncontroversial: native where interiors may matter, images where pages must be fixed, fields travelling either way. Made by default: whatever the platform emitted, whatever the last case did: it becomes the dispute that surfaces at the worst time: the flattened model whose formulas were the case, the "complete" production whose hidden sheets nobody produced, the authenticity challenge against a PDF whose native nobody can now find. This guide is the decision made properly, once, in the protocol.

The three families: native, searchable PDF and image-plus-load-file

- **Native:** the file as maintained, hash-verified, with its full interior: maximum evidence, minimum fixity: no stable pagination for citation, redaction impossible without conversion, and some formats needing their applications to read: the default, engineered around rather than abandoned.
- **Searchable PDF:** a rendered conversion carrying its own text layer: portable, paginated, bundle-ready: and a derivative: embedded metadata largely gone, dynamic content frozen as one moment's render, tracked changes shown or hidden per the converter's settings (a decision someone made, visibly or not).
- **Image + load file:** page images (TIFF or image PDF) with extracted/OCR text and the agreed metadata fields in structured companions: the litigation-industrial standard for redacted and page-controlled productions: interiors absent, fields preserved, guide 86's text-quality disciplines applying to the text layer.
- **Hybrids in practice:** real protocols mix: natives for interior-bearing classes, images for redacted items, PDFs for correspondence: per-class rules beating one-size answers.
- **Whatever the family, the constants:** hashes tying productions to collected originals (guide 82 §7's integrity loop), family relationships intact (guide 81 §5), and the conversion fact flagged where conversion happened (guide 87 §11's clause).

§ 0 4 · THE FORCING CLASSES

Document classes that force the choice: spreadsheets, presentations, chat and media

- **Spreadsheets and models:** the paradigm case: formulas, precedents, hidden sheets and columns, filters, pivot caches: the interior is the evidence, and imaging destroys it while multiplying page count absurdly: native always, absent express agreement, with guide 71's structured-data disciplines for the database-grade ones.
- **Word-processed documents:** tracked changes, comments and version internals matter exactly when drafting history matters (guide 75's territory): native where it might; PDF acceptable for the correspondence file.
- **Presentations:** speaker notes and hidden slides: the candid layer under the polished one: native where the deck's history or notes could count.
- **Chat and messaging:** no native "document" exists in the page sense: production is the unitised export per guides 63-66 and 81 §6, format questions becoming unitisation-and-fields questions: the protocol says what a produced conversation is.
- **Media and CAD/specialist formats:** audio, video, images and technical formats produce native with players or viewers per guide 77 §5's logic: a screenshot of a video is not the video.

The rules layer: PD 57AD, PD 31B and what "native" means in an order

- **PD 57AD's default:** disclosure in the form in which documents are ordinarily maintained, metadata intact unless agreed or ordered otherwise: the native-first posture written into the Business and Property Courts' regime, with the DRD the venue for the per-class detail.
- **PD 31B's parallel:** outside PD 57AD, the e-disclosure practice direction runs the same direction: native format expectations and metadata preservation: the drafting habit transfers.
- **Orders that say "native":** mean the maintained form with its interior: parties producing conversions against such orders are answering for it: Example 2's territory: and parties wanting imaging carve it out expressly, for cause.
- **Criminal and tribunal contexts:** disclosure regimes differ (CPIA's own rules; tribunals' informality) but the evidence logic does not: interiors matter wherever authenticity or process could: the format conversation had early in any forum.
- **The costs shadow:** format defects are re-production, re-review and application costs: courts allocating them to the party whose choice caused them: the drafting hour repaying itself in the usual currency.

Redaction, confidentiality and the imaging exception

- **Redaction forces fixity:** reliable redaction requires a fixed rendering (guide 93 treats irreversibility): the legitimate imaging case, applied per document, with the unredacted native preserved behind it.
- **Confidentiality regimes:** sealed schedules, confidentiality rings and external-eyes-only tiers often run on paginated images for control: cause stated, natives held for the ring's inner tier where interiors matter.
- **Flag the conversion:** per guide 87 §11: imaged-for-redaction documents identified as conversions, so the receiving party knows what it is not seeing and can invoke the route back.
- **Partial-native solutions:** spreadsheets redacted by producing native copies with redacted cells cleared and the operation logged: engineering that keeps the interior while meeting the confidentiality need: negotiated per class, documented per file.
- **Metadata in the redaction path:** redaction tools alter what files carry (guide 93 again): the production workflow verifies the redacted output's metadata posture rather than assuming it.

Receiving productions: acceptance checks and challenge routes

- **The intake audit:** formats received against formats agreed; hashes verified; load files reconciled (fields present, families linked per guide 81 §5); conversions flagged as flagged: the checklist run before review starts, not after it stumbles.
- **The flattening challenge:** spreadsheets or interior-bearing classes arriving imaged without agreed cause: challenged promptly with the protocol in hand: Example 1's letter, sent early, costs a paragraph; sent late, an application.
- **Text-quality inheritance:** received image productions carry the producer's OCR: scored per guide 86 §4 before your searches trust it, with re-OCR of their images your unilateral option.
- **The native request:** for any produced document whose interior or authenticity matters, the route back invoked: native original, provenance, platform records: guide 87 §11's machinery, exercised as routinely as it was drafted.
- **Proportional escalation:** format complaints ranked by materiality: the flattened model that is the case versus the PDF letter that is fine: courts rewarding the party whose challenges carry counts and consequences, per the standing arithmetic habit.

§ 0 8 · THE WIDER MAP

Source architecture: where else the evidence lives

Per this series' source-architecture discipline: a badly formatted production is a window, not a wall: what the produced form stripped usually survives elsewhere, and the receiving party's remedies are a map, not a grievance. The stripped interior may live in: the producing party's natives (preserved behind every conversion: the route back of §7, enforceable); your own client's counterpart copies (the same spreadsheet as exchanged in the deal, interior intact: the first place to look and the cheapest); the platform layers behind the documents (DMS ladders per guide 75, cloud versions per guides 51-58: server-side interiors no production format touched); the transmission record (the model as emailed, native in the attachment, dated by the servers per guide 53); structured sources upstream of the document (the ERP rows the spreadsheet summarised, per guide 74: often better evidence than the spreadsheet ever was); and the forensic layer (the producing custodian's device holding the file's local biography per guide 41). Format disputes shrink when the receiving party arrives holding the interior from another source: and the §7 challenge letter that encloses your counterpart native's differing content is the short version of this whole section.

WHAT THE PRODUCTION STRIPPED	PRODUCER'S NATIVES	YOUR COUNTERPART COPIES	DMS / CLOUD PLATFORMS	TRANSMISSION RECORD	UPSTREAM STRUCTURED SOURCES	DEVICE LAYER
Formulas & hidden content	Y: the route back	Y: as exchanged	Version interiors	Native attachments	The rows beneath	Local copies
Tracked changes / notes	Y	Drafts as circulated	Y: ladders	Y: drafts sent	n/a	Autosave strata
Embedded metadata	Y	Your copy's fields	Audit records	Header-dated copies	n/a	Filesystem story
Pagination-destroyed context	Native layout	Y	n/a	n/a	Re-query beats re-read	n/a
Provenance itself	Collection records	Independent existence proof	Y: server histories	Y: the dated lattice	Y	Y: guide 41

Fallback strategy in one line: before fighting about the format, fetch the interior from the estate: counterpart copies, platforms, transmissions and upstream systems usually hold what the production stripped: then fight, if still needed, holding it.

Worked examples

EXAMPLE 1 · THE MODEL PRODUCED AS 4,100 PAGES

A valuation dispute's central exhibit was the defendant's financial model: produced, despite a native-format protocol, as a 4,100-page PDF of every worksheet at print scale. The claimant's intake audit (§7) caught it in week one: the challenge letter cited the protocol, enclosed the page count, and noted that formulas, precedents and the model's hidden "Scenarios" sheet: visible in the PDF's bookmarks but absent from its pages: could not be produced by printing. The native arrived within fourteen days under threat of application. Its interior repaid the letter: the hidden sheet held the earlier valuation at 40% below the pleaded figure, and cell comments recorded who had ordered the headline changed. The PDF had not concealed these things maliciously: the producing provider's default imaging workflow had simply destroyed them on the way out: which is exactly why the protocol had said native, and why the intake audit exists.

EXAMPLE 2 · THE "NATIVE" ORDER ANSWERED WITH CONVERSIONS

An order required production of drafting-history documents "in native format with metadata intact". The production arrived as freshly generated PDFs: each Word draft converted the week of production, tracked changes accepted, comments gone, every file's embedded record showing the same seven-day creation window. The receiving party's expert report needed one table: order wording, format received, per-file producer strings and creation dates: to establish the breach; the §8 lattice then rebuilt what the conversions had stripped, pulling the true drafts from the claimant's own counterpart copies and the producing party's DMS ladder under a follow-up order. The costs of the exercise landed on the producing party, with the judgment noting that "native" is not a style preference but a description of evidence: the sentence this guide's §5 exists to make unnecessary.

EXAMPLE 3 · THE IMAGING THAT WAS EXACTLY RIGHT

A pharmaceutical dispute required production of batch records containing third-party confidential data and patient-adjacent identifiers. The protocol did what §6 describes: the class was imaged for redaction, the cause stated per document, redactions applied to the fixed renders under guide 93's disciplines, the agreed metadata fields travelling in the load file, and the unredacted natives preserved under a confidentiality ring's inner tier with an inspection mechanism. When one redacted table's underlying figures became genuinely material, the ring mechanism produced the native cells to external experts within days, without dispute. Nothing about the case's format story reached a judge: the measure of a protocol working: because the imaging was for stated cause, the fields travelled, and the route back had been drafted before anyone needed it.

Common mistakes and technical limitations

Common mistakes

- **Format by platform default:** the provider's standard workflow standing in for a decision nobody made: Example 1's cause.
- **Spreadsheets imaged:** the forcing class flattened: interiors destroyed, page counts absurd, the case sometimes inside.
- **"Native" orders answered with conversions:** Example 2's breach: fresh PDFs wearing the order's word.
- **Conversions unflagged:** the receiving party unaware it is reading derivatives: guide 87 §11's clause absent.
- **Load files unreconciled at intake:** missing fields and broken families discovered mid-review: §7's audit skipped.
- **Redaction natives discarded:** the unredacted originals not preserved behind the images: the route back demolished.
- **Chat produced as screenshots:** the guide 63-66 estates flattened to pictures: unitisation and fields both lost.
- **Late format challenges:** the flattening accepted silently for months, then raised at trial preparation: the remedy aged into an application.

Technical limitations

- **Natives need environments:** some formats require their applications or viewers: production includes the means to read, per guide 77 §5's playback logic.
- **Pagination and natives conflict structurally:** citation schemes for native productions (document IDs, cell references) are engineered, not assumed: bundles and natives coexist by design.
- **Dynamic content defeats fixity:** linked spreadsheets, live queries and auto-fields render differently per opening: material dynamics captured as values-plus-native pairs, stated in the protocol.
- **Conversion is lossy in both directions:** imaging strips interiors; but native-only productions of visually complex documents can lose the authoritative appearance (the signed page as signed): classes needing both get both.
- **Volume economics differ:** imaging multiplies storage and page-rate costs on large sets: proportionality arguments run on real numbers per the standing arithmetic habit.

§ 1 1 · INTERROGATORIES & DRAFTING AIDS

Questions to ask · Suggested wording

Ask your client

- Which document classes in this matter carry interiors that could count: models, drafts, decks: for the native list?
- What confidentiality constraints will force imaging: and is the ring-and-route-back machinery drafted for them?
- Can our side actually open what we are demanding natively: viewers, licences, specialist formats?

Ask your opponent

- Please confirm the format in which each class of documents will be produced, identifying any class to be produced otherwise than in native format and the cause in each case.
- Please identify any produced document that is a conversion from native format, and confirm that unredacted native originals are preserved and will be made available under the agreed mechanism where authenticity or non-visible content is in issue.
- Please confirm that spreadsheets, presentations and other interior-bearing documents will be produced in native format with hashes, and state the citation convention proposed for native documents.

Ask your eDiscovery / forensic provider

- Does the intake audit reconcile: formats, hashes, fields, families: before review begins?
- Which received documents are conversions, and which merit the native request now rather than later?
- For our own production: is the workflow honouring the protocol per class, and flagging every conversion it makes?

SUGGESTED WORDING · FORMAT PARAGRAPH FOR THE DISCLOSURE PROTOCOL

"Documents shall be produced in native format, being the form in which they are ordinarily maintained, with hashes and the metadata fields at Appendix [A], save that: (a) documents requiring redaction, and documents within the classes listed at Appendix [B] for the causes there stated, may be produced as searchable page images with extracted text and the Appendix [A] fields; (b) any document produced otherwise than in native format shall be identified as a conversion; and (c) spreadsheets, presentations, databases and other documents whose non-visible content may be material shall be produced in native format absent express written agreement. Unredacted native originals of all converted or redacted documents shall be preserved, and shall be made available in accordance with paragraph [X] where the authenticity, provenance or non-visible content of a document is in issue. Native documents shall be cited by document identifier and, where applicable, internal reference (sheet and cell, slide, paragraph)."

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

The format checklist

- ☐ Per-class format table agreed at the DRD stage
- ☐ Native default; imaging causes stated per class or document
- ☐ Spreadsheets, decks and databases on the always-native list
- ☐ Conversions flagged; unredacted natives preserved behind them
- ☐ Load-file fields and hashes travelling with every family
- ☐ Citation convention for natives agreed before bundling
- ☐ Intake audit run on every received production
- ☐ Received OCR scored before searches trust it
- ☐ Native requests exercised early for contested documents
- ☐ Format challenges made promptly, with materiality and numbers

Red flags

- Spreadsheets arriving as page images: Example 1's shape.
- Productions under "native" orders with same-week producer strings: Example 2's table.
- Conversions nowhere flagged as conversions.
- Load files missing agreed fields, families unlinked.
- Redacted productions whose unredacted natives "were not retained".
- Chat evidence produced as screenshots.
- Format objections raised for the first time at trial preparation: by either side.

When to involve a digital forensic expert

At protocol drafting, where the per-class table, the imaging causes and the route back are engineered (Example 3's uneventful success); at intake, where the audit catches Example 1's flattening in week one; at challenge, where conversion breaches are evidenced (Example 2's producer-string table) and stripped interiors are rebuilt from the §8 lattice; and at production, where your own workflow honours the protocol and flags what it converts. Under CPR Part 35, format and provenance findings report with the guide 87 disciplines behind them. Through **e-discovery.uk**, productions run native-by-default with the machinery of this guide as standard practice.

§ 13 · COMMON QUESTIONS

Frequently asked questions

Is there a single best production format?

Per class, yes; universally, no: native where interiors may matter (which is more classes than intuition suggests), images where redaction or confidentiality forces fixity, fields and hashes travelling regardless: §3's hybrid is what competent protocols actually say. The one universally wrong answer is the undecided default.

The other side wants everything as PDFs "for convenience". Should we agree?

For correspondence-grade classes, often harmless; for interior-bearing classes, it is an evidence concession dressed as logistics: Example 1 is what "convenience" can cost. Agree a per-class table instead: PDFs where interiors cannot matter, natives where they can, and the route back for surprises: convenience and evidence both survive.

How do we cite a native spreadsheet in a bundle or judgment?

By identifier and internal reference: document ID plus sheet and cell: with key extracts rendered to fixed pages for the bundle while the native remains the evidence: §10's engineering point. Courts handle this daily; the failure mode is not citation difficulty but nobody agreeing the convention until the bundle deadline.

We received images and the text layer seems poor. Whose problem is that?

Practically yours until challenged: your searches inherit their OCR: so the guide 86 response runs: score what arrived, re-OCR their images yourself where it pays, and put quality questions per §11 where the production falls short of the protocol. The protocol drafted with guide 86's paragraph makes this a compliance issue rather than a philosophical one.

Do we have to produce hidden sheets, tracked changes and speaker notes?

If the document is disclosable, its non-visible content is part of it: producing a version with the interior stripped is producing a different document: Example 2's principle. Privileged or irrelevant-and-confidential interior content is handled by redaction with the conversion flagged: not by silent flattening.

The producing party says the native no longer exists. Now what?

Test it, then route around it: preservation duties make "not retained" a serious statement (when did it exist, what happened, per the spoliation disciplines): while the §8 lattice rebuilds the interior from counterpart copies, platform versions, transmissions and upstream systems: Example 2's rebuild. The absence becomes its own chapter; the content usually turns up anyway.

§ 1 4 · REFERENCE

Glossary

Conversion

A produced derivative of a native file; flagged as such.

Fixity

The stable, paginated form imaging provides.

Hybrid protocol

Per-class format rules mixing native and imaged production.

Image + load file

Page images with text and metadata companions.

Interior

A document's non-visible content: formulas, tracks, notes, metadata.

Native format

The file as ordinarily maintained, interior intact.

Producer string

The PDF's embedded record of what created it (Example 2's witness).

Ring (confidentiality)

The tiered-access regime imaging often serves.

Route back

The preserved right to natives and provenance on request.

Values-plus-native pair

A captured render beside the dynamic native.

Sources and authoritative references

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

- e-discovery.uk, EDRM Phase 06 · Production (format engineering, load files and intake audits as practised): e-discovery.uk/edrm/production · Phase 04 · Processing: e-discovery.uk/edrm/processing
- cflab.uk, digital forensics services (provenance examination, conversion-breach evidence, interior reconstruction, CPR Part 35 reporting): cflab.uk/digital-forensics-services · guides library: cflab.uk/guides
- e-discovery.uk, practice method and Knowledge Centre: e-discovery.uk · e-discovery.uk/knowledge
- PD 57AD (form of production and the Disclosure Review Document): justice.gov.uk, PD 57AD · PD 31B: justice.gov.uk, PD 31B
- CPR Part 35: justice.gov.uk, Part 35 · ISO/IEC 27037: iso.org, 27037
- ICO, UK GDPR guidance: ico.org.uk

DISCLAIMER

This publication provides general information about production formats as at August 2026. Platform behaviour and format capabilities vary by product and version; verify the current position for the tools in the matter at hand. The worked examples are fictional. This guide is not legal advice, does not create a professional relationship, and should not be relied upon in place of advice on the facts of a specific matter from a qualified lawyer or a suitably qualified forensic practitioner. Links were live at the date of publication.

§ HOW A SPECIALIST LABORATORY CAN ASSIST

Working with Computer Forensics Lab

Format work at **e-discovery.uk** runs both directions of this guide: productions engineered native-by-default with per-class causes, flagged conversions, travelling fields and preserved routes back (Example 3's uneventfulness as the goal); and intake audits that catch the Example 1 flattening in week one, score inherited text layers per guide 86, and exercise the native request while it is still a letter. The laboratory side evidences the breaches (Example 2's producer-string table), rebuilds stripped interiors from the §8 lattice, and reports provenance under CPR Part 35 where format has become the fight. Conflicts checks, confidential scoping discussions and fixed-fee estimates are routine; and if a spreadsheet that matters has just arrived in your production as four thousand pages of PDF, the challenge letter drafts itself: this week, while it is cheap.



I N S T R U C T T H E L A B

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