

LEGAL-GRADE PDF REDACTION

Redacted content must be irrecoverable, not just invisible.

In legal workflows, redaction must ensure complete removal of sensitive content—text, metadata, annotations, and embedded objects—not just visual masking. Incomplete redaction may expose privileged or personal data, risking GDPR violations and legal liability. Accurate, standards-based redaction is critical for maintaining compliance and protecting confidentiality.

ANNOTATIONS

n of PDF/UA conformance places strings
n both document authoring software and

To be considered:
Remove whole annotation?
Flatten annotation?
Remove partially?
Based on annotation type?

PDF/UA
of PDF/UA conformance places strings in both document authoring software and the PDF text strings of arbitrary lengths may occur

Apply eraser tool in Ink annotation?
Remove part of the annotation appearance?
Make sure no other data enables it's reconstruction.

EDGE CASES

Text 1
Text 2
Text 3

Incremental save
Image metadata
Font metadata
OCR
Redacting repeated
Images Form XObjects

Duplicate objects.

Remove data from
all object occur-
rences.

Text 1
Text 2
Text 3

IMAGES

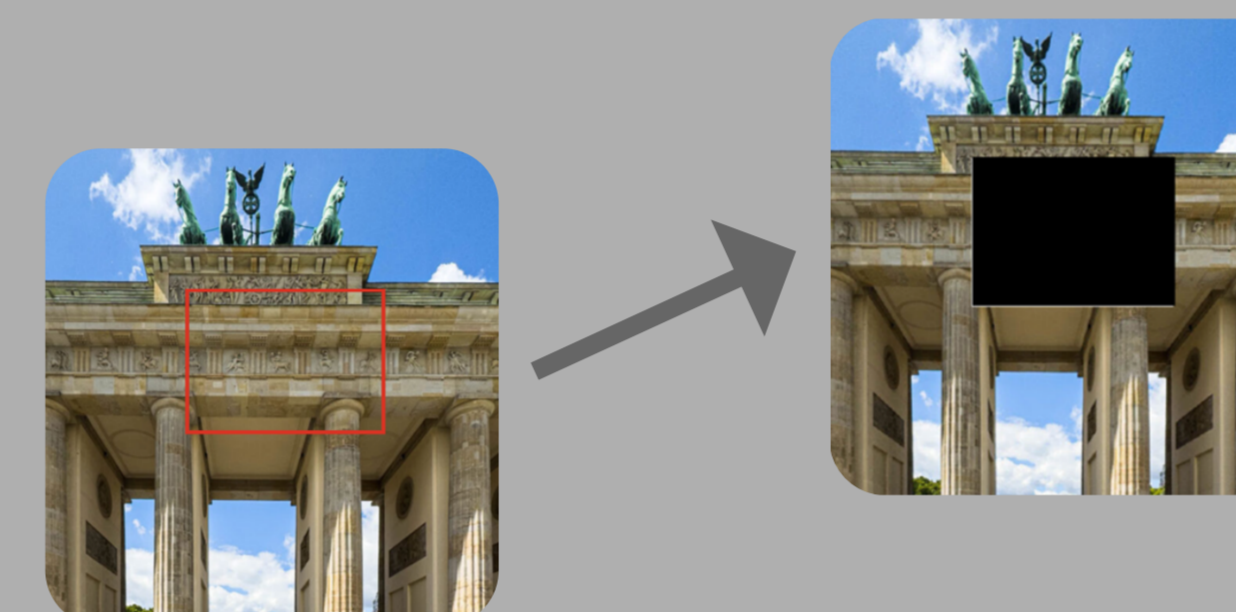


Image data need to be changed
Images support formats: JPEG, JBIG2,
JPEG2000, TIFF.
Replace or re-encode image data
avoid artifacts or size inflation.
Redact image metadata
Consider clip
Image mask

HIDDEN OBJECTS

Metadata
Embedded content and attached files
Scripts
Hidden layers
Embedded search indexes
Stored form data
Review and comment data
Tags and marked content
Comments hidden within the body of the PDF file
Hidden data from the previous document saves
Obscured text and images
Unreferenced data
Links, actions, and javascripts
Overlapping objects

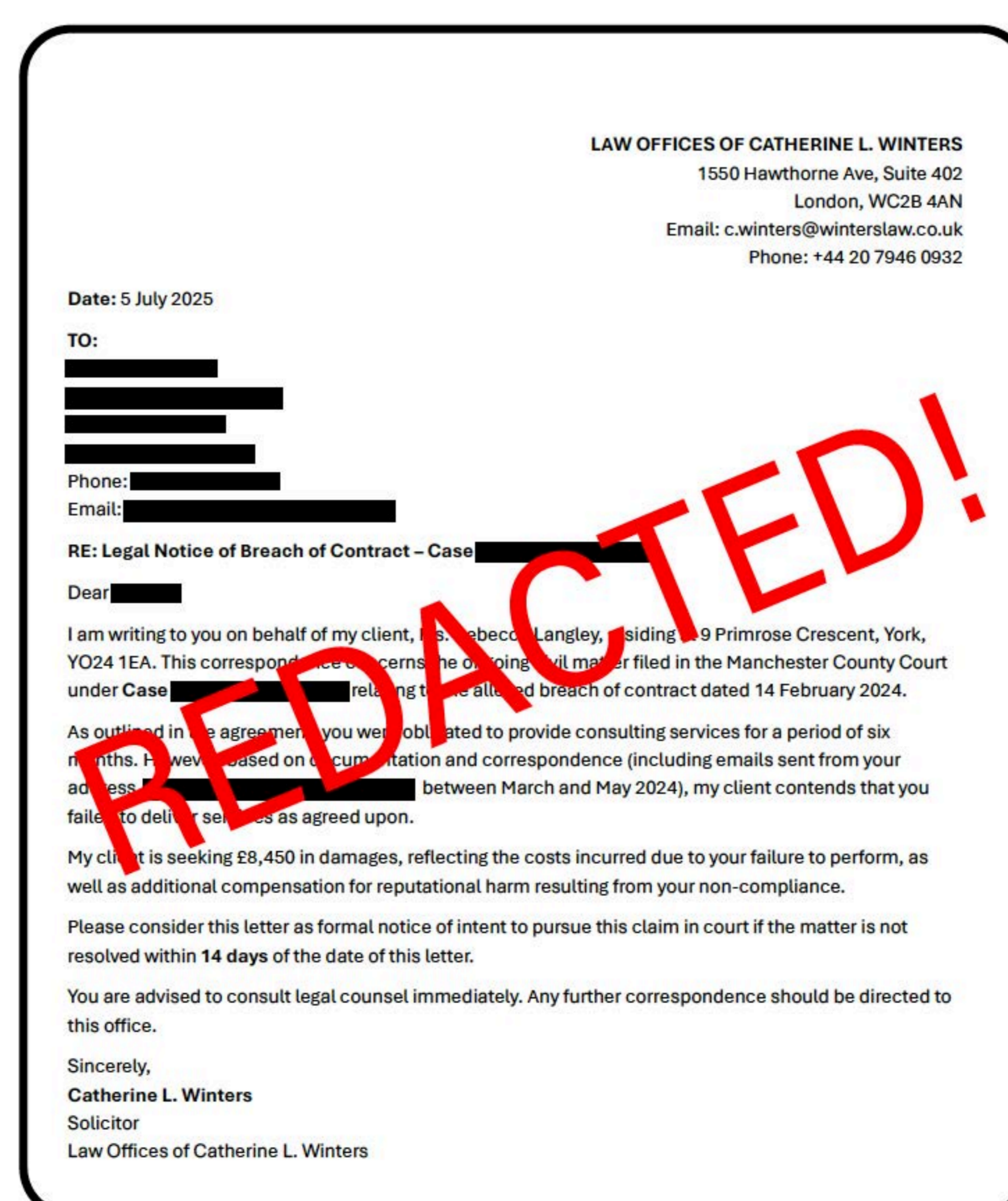
TEXT

The exact length of the removed text
also remains recognizable, which may
help in guessing plausible wordings for
shorter redacted sections. Where
computer-generated proportional fonts
were used, even more information can
leak out of the redacted section in the
form of the exact position of nearby
visible characters.

Text selection is undefined
Search doesn't work
Text styling
Rotated text
Overlapping objects

The exact length of the removed text
also remains recognizable, which may
help in guessing plausible wordings for
shorter redacted sections. Where
computer-generated proportional fonts
were used, even more information can
leak out of the redacted section in the
form of the exact position of nearby
visible characters.

Partial intersection
Hidden text
Text clipping
Remove unused characters
Remove font metrics



PATHS & SHAPES



Original vector object



Dash line changed



Converted to bitmap

Paths consist of:
Geometry
Stroke and Fill
Line ending
Line dash
Line style

To be considered:
Conversion to bitmaps
Patterns may contain repeated content
Shading patterns contains computed contents
Clipped pattern
Remove whole pattern