

From Chaos to Clarity

The Forensic Data Processing Pipeline: Transforming massive volumes of raw data into a manageable, reviewable, and defensible evidence set.

The Processing Pipeline

A non-negotiable sequence to normalize, enrich, and filter raw data, ensuring every file is treated equally and irrelevant information is removed before review begins.

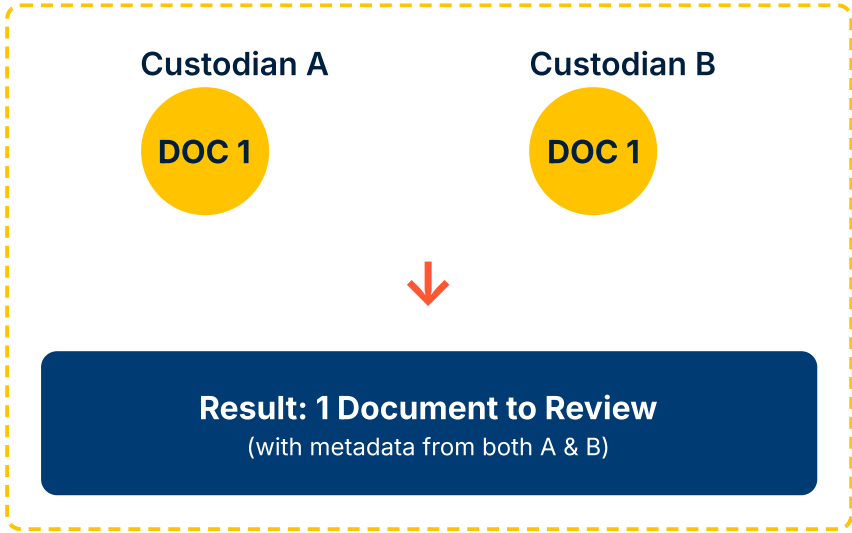


Strategic De-Duplication

The choice between Global and Custodial de-duplication is a critical strategic decision that impacts review cost and defensibility. It is based solely on a file's cryptographic hash (e.g., SHA-256).

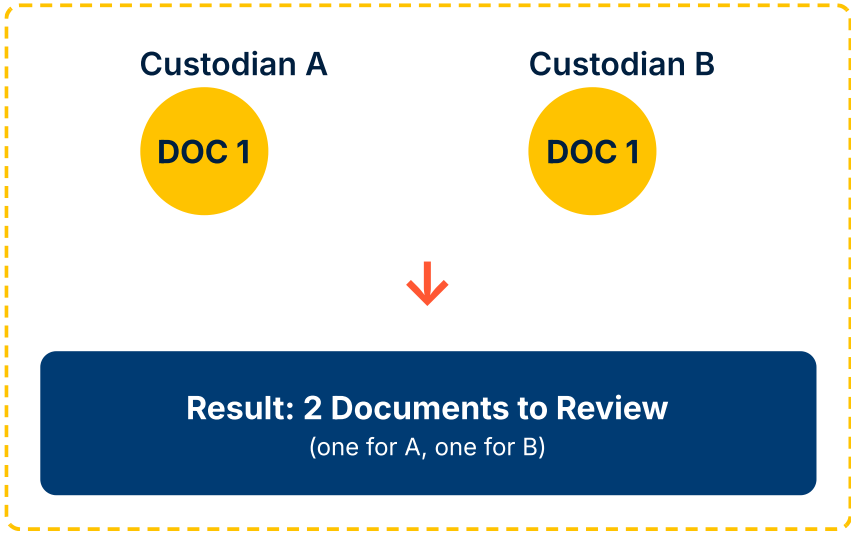
Global De-Duplication

Only one unique copy of a file is reviewed, regardless of how many custodians had it. This is cost-effective but requires careful metadata preservation.



Custodial De-Duplication

A file is only removed if it's a duplicate within the *same* custodian's data. This increases review volume but clearly proves possession.



Typical Document Reduction Funnel

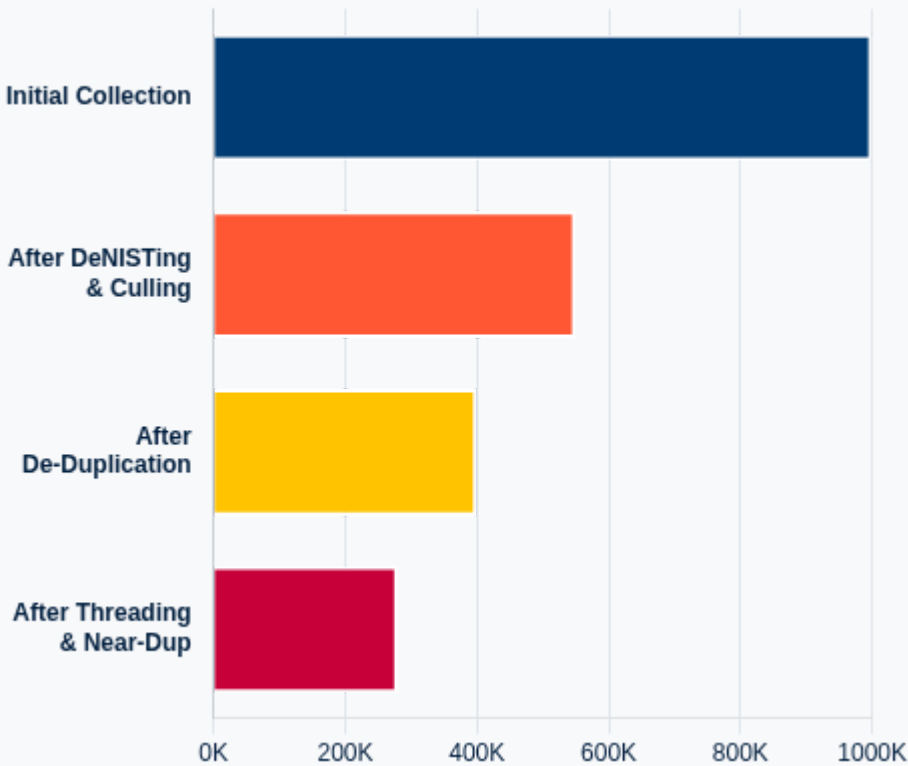
The Impact: Drastic Data Reduction

Advanced processing techniques dramatically reduce the volume of data requiring expensive human review. The chart below shows a typical reduction funnel, where an initial collection of 1 Million documents is culled by over 70%.

DeNISTing & Culling: Removes irrelevant system files and applies date/keyword filters.

De-Duplication: Removes exact duplicate files based on hash values.

Threading & Near-Dup: Suppresses redundant emails and groups similar document drafts.



Advanced Efficiency Techniques

Beyond exact duplicates, these methods intelligently group related documents to accelerate the review process.

