



Obliterating an eDiscovery log jam with Intella Desktop

Use Case

I recently became involved in a *pro bono* matter, with attorneys attempting to provide assistance to a small non-profit in responding to an onerous government subpoena. The subpoena threatened the continued existence of the organization.

There was far too much data to be reviewed, the data repositories, types and volume were largely unknown, the six-figure estimates for outside assistance were completely out of the question, and no one saw a path forward.

After performing some basic initial assessment, I quickly determined that, from an eDiscovery perspective, although the situation was

indeed serious, none of the problems were as insurmountable as described.

Although situations like this occur every day in eDiscovery, the only permissible “solution” can ever be the one-size-fits-all model, primarily designed to support the world’s largest corporations and law firms, via the multi-billion dollar eDiscovery service provider industry.

What was needed here is technology that is powerful, but also affordable, and that empowers practitioners to solve a pressing problem, without facing insolvency as the only possible alternative.

Like many, this was a project made for Intella.

Collection

In reality, most of the potentially relevant email data had already been preserved and was available in PST format, with a small additional volume located in a common cloud repository.

The latter was collected first, and processed directly into Intella. Within minutes, content that seemed abstract and was poorly understood for many months was instantly available, visualized, fully searchable, and was immediately subjected to further assessment via Intella’s vast, intuitive feature set.

The PST email was quickly obtained and processed into the same Intella case, resulting in federated searching of the full corpus of potentially responsive ESI.

Processing

The majority of the dataset, consisting of approximately 190 GB of PST content across a dozen or so custodians, was processed in **less than 90 minutes**. This performance was achieved on a modestly-priced (approximately \$1,000 USD) Dell workstation, featuring a large PCIe system drive, secondary HDD, 64 GB RAM, and a 10th generation Intel i7 Octa-Core processor.

Although relatively inexpensive as hardware goes, by adhering to Vound’s recommended hardware considerations, along with an understanding of how Intella makes use of system resources, the level of processing performance achieved far eclipsed prior experiences involving high-end processing setups from just a few years ago.

The processing performance aspects of Intella cannot be

overstated in a world of ever-expanding data volumes and constantly-evolving, complex data types. This is further supported by Vound’s development team, which works continuously to improve processing efficiency, accuracy and performance.

With the full dataset now available, several of Intella’s post-processing options were leveraged, including OCR, email threading, and (eventually) near-duplicate identification. It is noteworthy that Intella handled threading of over 1 million email items, which was applied proactively, both in anticipation of the presence of voluminous such threads, and in order maximize technology-based volume reduction techniques.

ECA Effort

In parallel with data handling, preliminary search criteria was developed based on the subpoena requests, and compiled in what eventually became a series of sequentially-named text files. This approach permits easy documentation of the iterative process that is the hallmark of effective, defensible ECA in Intella. Whenever significant changes occur from one iteration to the next, Save As to a new file, and continue the process. Adding dates to these files names (i.e. 001a primary KWs (20240510).txt) further facilitates efficient tracking both inside Intella and out, when new keyword lists are imported.

Of course, the heart of the iterative process in achieving effective ECA requires actually looking at your results. What’s poorly understood (or perhaps even verboten?) is the idea that quick spot-checking of results (via Intella’s industry-best

table view) can yield new, immediately-actionable information that prompts further iteration and removal of obviously irrelevant content that is not worthy of expensive attorney review time. This isn't "review" and needn't involve a random sample. Instead, it involves the long-practiced use of complex search criteria, iteratively modified to more effectively exclude obviously irrelevant content, in conjunction with the use of Intella's multiple facets, which provide endless angles via which to assess and evaluate your dataset.

The Result

That all sounds great, but what does this look like where the rubber meets the road?

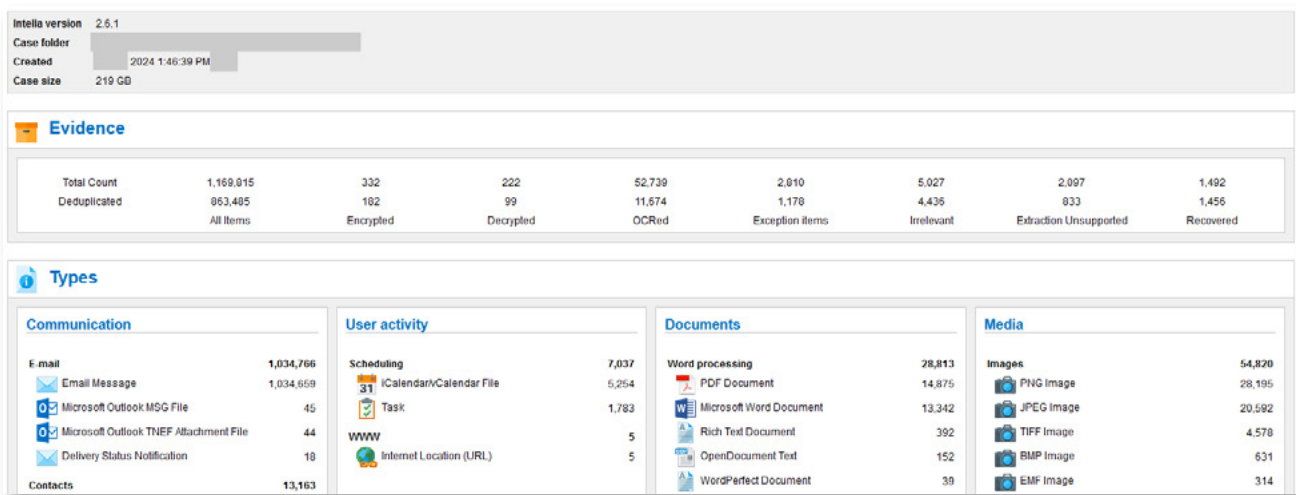
In this case, it meant a reduction of data volume from 1,169,815 total items, to 863,485 with de-duplication applied; further reduction to 11,302 items after iterative keyword searching (11 total iterations); and additional reduction via extensive exclusion

filtering and culling to 7,557 items (including families) for load file export and eventual attorney review.

Below is a screenshot from the Intella Insight report from the actual case:

Even if evaluating most conservatively against the 863K unique items, that's a >99% cull rate, or a resulting review set that is .87% of the initial volume. Also, this is not an aberration, as I have achieved similar results in exponentially larger data sets, and have done so in over 500 cases of almost every possible type within the realm of civil eDiscovery. Across all of those matters, I would estimate an anecdotal post-cull volume of between 1 and 3%, with something like 5% usually indicative of an incomplete or non-optimized ECA effort.

Obviously, these are generalizations, and every matter is unique, but this is what I have experienced over 15 years of sustained Intella usage.



Project Success Factors

Tagging is key

As content is suppressed, it's critically important that tags be applied that accurately track the work that was done in order to arrive at your results. This supports defensibility of process.

Standardizing the tags and tag groups used across all of your matters will further increase accuracy and efficiency.

Save your searches

This applies not only to keyword queries, but exclusions, as well. In my case, that list became so complex that it was saved totally by itself, such that it could be retrieved first, and then the revised by adding the current iteration keyword list as a secondary step. Then that iteration of the search result could be saved.

Save your reporting

Take advantage of Intella's industry-best keyword statistics reports (I prefer the PDF format), appropriately named, to further bolster defensibility of process, and serve as snapshots in time of the iterative ECA effort.

Back up your data

Computer components fail, and days or weeks of very expensive work product can be lost as a result.

Back up your data at regular intervals, and pursue solutions that offer both local and cloud options, placing the former on separate media. If a hard deadline falls on the following day, it matters little that a cloud service can provide a restoration disk in a week, or that 500 GB can be downloaded in 24-36 hours.

Best of luck to you in improving your work with Intella!



Phone Enquiries
+1 (888) 291-7201

Postal Address
10643 N Frank Lloyd Wright Blvd, Suite
101, Scottsdale, AZ 85259 U.S.A.

Email sales@vound-software.com
Sales Contacts
www.vound-software.com/partners

