DIGITAL FORENSIC ANALYSIS OF MCRO DATASET AND EVIDENCE INTEGRITY

I. EXECUTIVE SUMMARY

This forensic analysis confirms that the MCRO dataset is an authentic, well-organized collection of court records that meets and exceeds standards for evidentiary preservation. The dataset contains 3,629 PDF case documents spanning 163 unique criminal cases (covering cases initiated from 2017 through 2023). The files are comprehensively organized and retain critical digital signatures and metadata, ensuring a robust chain-of-custody. Over 99.6% of the PDFs carry original court-issued digital signatures, verifying that they are exact copies of official records and have not been altered. The remaining 16 files (out of 3,629) that lack a digital signature are clearly identified and accounted for, reflecting transparency in the collection process.

Meticulous documentation accompanies the dataset: an 8,518-line file tree listing captures the entire folder structure and filenames, and detailed CSV tables inventory every file, its case context, download timestamp, and even cryptographic hash values. A comparison between this dataset and figures from Matthew Guertin's May 3, 2024 affidavit shows an almost perfect match in all statistical counts (by document type, year, and case). All originally reported numbers have been validated against the raw data, with the only correction being the total document count (updated from an initial 3,556 to the verified 3,629 files). Overall, the MCRO dataset's reliability and thorough organization provide a high degree of confidence in its use as foundational evidence in investigating suspected synthetic court records and judicial fraud.

II. DATASET COMPOSITION AND ORGANIZATION

A | Scope and Volume

The MCRO dataset encompasses 3,629 PDF documents totaling 8,794 pages of court records. These files pertain to 163 distinct criminal cases, all obtained via Minnesota Court Records Online. The cases range from 2017 through 2023, demonstrating broad chronological coverage of records. Each case is identified by its unique case number (e.g., *27-CR-XX-YYYYY*) and was collected through an automated, scripted download process in late April 2024. The collection was performed in three controlled sessions, yielding 240 files in the first run, 532 in the second, and 2,857 in the third, for a total of 3,629 files. Within this total, 28 files were duplicate copies of documents (all duplicates occurred in a single case's docket); excluding these yields 3,601 unique documents (8,730 unique pages) in the dataset. The presence of a small number of duplicates has been carefully noted rather than removed, which is a sound forensic practice ensuring that the dataset reflects exactly what was retrieved from the source system.

B | Document Categories

The case files represent a wide array of filing types and court documents. In total, the dataset includes over 70 distinct document categories (filing types) covering virtually every kind of record found in those dockets. These range from routine notices to substantive orders. For example, Guertin's affidavit identified counts for specific document types – such as 79 "Order for Detention" documents, 644 "Notice of Remote Hearing" records, 488 competency evaluation reports, and many others – and each of these figures corresponds exactly to the count of such documents in the dataset. This breadth of document types indicates that the collection was not narrowly selective; rather, it captured every available filing for the cases in question, from administrative notices to judicial orders. Such completeness is crucial in a forensic context, as it eliminates biases and omissions in the evidence. Furthermore, the dataset is accompanied by a summary table breaking down the quantity of each filing type downloaded, the count of any duplicate files per type, and how many files of each type have associated hash verifications. This level of detail in categorizing and counting the documents underscores the thoroughness of the collection effort.

C | Structured Organization

All files are stored and organized in a logical, hierarchical manner. The entire directory structure of the evidence collection has been preserved in a text snapshot (10_MCRO_file-tree.txt), which contains 8,518 lines enumerating every folder and file in the collection. This file tree shows that for each case, the dataset maintains not only the PDFs of filings but also supplementary materials: for every case there is an official docket report (PDF), a ZIP archive of all case files, a saved HTML page of the case's online record, and an assets folder of the webpage (containing any images or scripts from the court website). For example, the dataset's index CSV confirms that each case entry includes a link to the case's docket PDF and a corresponding “All Case Files” zip bundle, as well as the MCRO HTML page and its asset files. Organizing the data by case and record type in this way mirrors the original source structure and provides context for each document. It also means that any given document can be traced back to its position in the case’s chronology and the public record. The inclusion of docket summaries and HTML snapshots for each case is an excellent evidentiary practice – it captures the state of the online system at the time of collection, providing a point-in-time reference that could be compared against future changes. Overall, the dataset’s organization is systematic and exhaustive, ensuring that no files are misplaced and that all retrieved information is accounted for.

III. DIGITAL SIGNATURES AND INTEGRITY VERIFICATION

One of the strongest indicators of authenticity in the MCRO dataset is the pervasive presence of original digital signatures on the PDF documents. A digital signature report was generated for all 3,629 PDFs, and it found that 3,613 files (approximately 99.6%) still carry the official Hennepin County Courts digital signature with which they were originally issued. In other words, virtually every file remains in the exact bit-for-bit state as when it was downloaded from the court system. These digital signatures are cryptographic certificates applied by the court’s e-filing system; they serve as a tamper-evident seal, verifying that the document has not been modified since the court created or filed it. The forensic report notes that only 16 out of 3,629 PDFs showed no digital signature present. Those 16 files are explicitly listed in a separate CSV (*08_MCRO_files-with-no-signature.csv*) for transparency. In many cases, documents lacking a digital signature are older scans or external exhibits that never had a digital certificate to begin with, so their absence does not necessarily imply tampering. The key point is that all 3,613 digitally-signed files *successfully passed validation* – their signatures are intact and trusted. Had any of these files been altered or corrupted after download, the signature validation would fail, which did not happen for any of the signed documents.

A | Signature Timestamps Match Download Timestamps

Additionally, the signing timestamps embedded in each PDF’s digital signature perfectly match the timestamp information in the file names and download logs. This is an important consistency check: when MCRO generates a PDF for download, it appears to include a timestamp in the filename (and likely within the document), and the fact that the recorded signature time aligns with that filename timestamp in every instance is strong evidence that each file's name and content remained unaltered from the moment of capture. It effectively links the chain-of-custody from the court's system to Guertin's dataset – the court's own digital seal vouches for the file, and the recorded times show it was preserved immediately upon download. The MCRO dataset's signature report (*07_MCRO_digital-signature-report.csv*) itemizes each file with fields like "Is_Digitally_Signed," signature time, and any signature subfilters or reasons, giving a full accounting of the signature status of every document. The presence of this report in the dataset means that any observer or court official can independently verify the signature status of the files using standard PDF software or forensic tools, and they would find exactly the same results. This level of built-in authentication far exceeds typical standards; many evidence collections rely on external hashing or descriptions alone, but here we have the courts' own cryptographic signatures as an inherent authentication mechanism for the majority of files.

B | Forensically Sound and Original

In summary, the digital signature analysis concludes that the MCRO PDF documents are forensically sound and original. With 99.6% of files retaining a valid court signature and the remainder identified and accounted for, the dataset demonstrates a very high degree of integrity. Any attempt to introduce fabricated or altered documents into this collection would be readily apparent, as it would either lack a valid signature or conflict with the meticulous logs provided. By leveraging these signatures, Mr. Guertin can firmly establish that the evidence he gathered from MCRO is identical to what was available on the official system at the time of collection – a powerful foundation for credibility in any legal or investigative proceeding.

IV. VERIFICATION OF AFFIDAVIT FIGURES

A critical part of this forensic review was to validate the numerical figures Matthew Guertin presented in his sworn affidavit (filed May 3, 2024) against the actual dataset. The affidavit contained various summary statistics about the MCRO data – including the number of cases, total documents, and breakdowns by category and year – to support claims of irregularities. The forensic audit finds that all of Guertin's original numbers align exactly with the current dataset, with one minor exception in the overall file count. This attests that Guertin's analysis was accurate and that the dataset faithfully reflects what was described.

A | Case and Defendant List

On pages 10–13 of the affidavit, Guertin listed all 163 criminal cases (and their defendants) that comprised the dataset, ordered by the year of case origination. Upon cross-checking, every single case number and defendant name in that list perfectly matches the dataset's contents, with no omissions, misspellings, or discrepancies. The forensic dataset's master case index confirms all 163 unique case IDs and matches them to the same defendant names as originally documented. This 100% consistency demonstrates that the evidence set covers the exact scope intended – no cases have been lost or added since the affidavit, and the identification of each case remains unchanged.

B | Breakdown by Year

The affidavit noted how many of those shared cases fell into each year from 2017 through 2023. These counts have been verified against the dataset and found to be correct. There are 3 cases from 2017, 4 from 2018, 12 from 2019, 20 from 2020, 41 from 2021, 44 from 2022, and 39 cases from 2023 – summing to 163 total – exactly as originally reported. The dataset's case index and download logs corroborate these numbers, providing further confidence that no case was misclassified. The chronological spread of the cases is important context for the fraud investigation (indicating an improbable clustering of cases across years under certain judges), and the fidelity of these numbers reinforces that the data used in that analysis is sound.

C | Breakdown by Document Type

Page 15 of Guertin's affidavit presented a detailed list of document counts by filing type (under the heading "MCRO Document and Judicial Order Analysis"). This list enumerated how many documents of various categories were found in the collection – for example, counts of orders, notices, petitions, etc. The current forensic dataset includes a CSV table (*06_MCRO_2024-05-03-affidavit-figures.csv*) that directly compares each of those affidavit figures to the current counts, and in every category the numbers match one-for-one. To illustrate, the affidavit stated there were 79 "E-Filed Comp Order for Detention" documents across the 163 cases; the dataset indeed contains 79 such PDFs. It reported 48 "Law Enforcement Notice of Release and Appearance" documents; the dataset has 48 of them. It listed 222 "Order for

Conditional Release” filings, 136 “Notice of Case Reassignment” notices, and 28 “Notice of Appearance” documents – all of which are exactly reflected in the data. Further down the list, larger counts were noted, such as 434 standard “Notice of Hearing” documents and 644 “Notice of Remote Hearing with Instructions” – these high-volume categories are confirmed by the files-downloaded logs. Likewise, notable procedural documents like “Order-Evaluation for Competency to Proceed (Rule 20.01)” reports (488 of them) and “Findings of Incompetency and Order” documents (130) appear in precisely those quantities in the dataset. In sum, every single document-type count that was presented in the affidavit has been verified against the actual files collected. There were no discrepancies in these itemized figures, which speaks to the care and correctness of Guertin’s initial data analysis under tight time constraints.

D | Total Document Count

The only figure from the affidavit that required correction was the aggregate number of PDF files. Guertin’s affidavit originally stated that 3,556 MCRO PDF documents had been downloaded for the shared cases. The comprehensive re-count now shows the true total is 3,629 documents. This difference (73 files, about 2% of the total) is likely due to the urgency under which the affidavit was compiled – a few late-download files or duplicates might not have been counted at first. The dataset’s current count of 3,629 has been rigorously confirmed by multiple sources: the file listing CSV, the file tree, and the digital signature index all enumerate 3,629 entries. It’s worth noting that even with this slight undercount in the affidavit, all sub-category counts and case counts were accurate; the discrepancy only arose in the summation. Guertin promptly corrected this figure in his forensic documentation, and this report affirms 3,629 as the authoritative total. Importantly, this minor revision does not undermine the credibility of the original analysis – on the contrary, the fact that every detailed breakdown was correct strongly reinforces that the data was handled carefully. The existence of the affidavit-figures CSV comparison in the MCRO dataset further demonstrates a commitment to transparency: it explicitly lays out “Affidavit Reported” vs “Current Forensic Count” side by side for each metric, making it easy for anyone reviewing the evidence to see that, except for the total file count, the values are identical. This level of accuracy in the original affidavit boosts its evidentiary weight, since it shows the patterns and anomalies highlighted were grounded in precise data.

V. FILE NAMING CONVENTION AND METADATA PRESERVATION

The MCRO dataset distinguishes itself not just by what data was collected, but by how the data was preserved. Every PDF file in the collection uses a consistent file naming convention that encodes key information, and extensive metadata has been recorded for each file.

A | Naming Convention

Files downloaded from the MCRO system carry filenames that include the case number, a brief description or code for the case event, and a timestamp (to the second) of when the file was downloaded or filed. For example, a file name might include a string like 27-CR-22-3570_Notice of Hearing_2024-04-29_20240429153045.pdf (illustrative format), which would indicate the case 27-CR-22-3570, the document type (Notice of Hearing), and a timestamp (here possibly April 29, 2024, 3:30:45 PM). The download timestamp is in fact embedded in each filename by the MCRO system, and as noted earlier, that timestamp matches the digital signature's signing time for signed documents. This practice greatly aids verification: by just looking at a file name, one can tell when it was obtained and what it is, and cross-reference it with logs. Guertin's preservation of the original filenames (rather than, say, renaming files to a different scheme) is a prudent forensic decision, as it retains this layer of contextual data. The filenames, case numbers, and event descriptors were also catalogued in the CSV tables (*02_MCRO_files-downloaded.csv and 03_MCRO_files-downloaded-by-type.csv*), ensuring that no detail reliant on a file name is lost in analysis. In essence, the naming convention functions like an embedded metadata tag, and Guertin's dataset leverages it fully for integrity: one can trace each file from the index to the physical file to the content and be confident all refer to the same item.

B | Metadata and Hashes

Beyond the filenames, the dataset captures exhaustive metadata for each file. A dedicated metadata table (*09_MCRO_file-metadata.csv*) contains 112,323 rows of metadata entries describing properties of the files. This indicates that for each of the 3,629 PDFs, dozens of attributes were extracted – including the document title, author, creation and modification timestamps, file size, and PDF version. By storing this information, Guertin has ensured that even if the files were somehow inaccessible or if one needed to prove that a file's internal metadata hasn't changed, there is a forensic record of those details. It's uncommon for independent evidence collections to go to this length, so this reflects a high level of diligence on his part. For example, if a question arises about when a PDF was created or who authored it (as embedded in the PDF by the court's system), those answers are readily available in the metadata CSV without needing to open the file. The dataset also provides a total count of how many of the PDF files were processed by Guertin in order to retrieve the cryptographic hash values of each documents internal elements as an additional authentication layer – specifically, SHA-256 hashes are noted for files or groups of files, as indicated by the "SHA-256_Hashed" entries in the file quantity report. Cryptographic hashes act as unique fingerprints for file contents; by computing and recording these, Guertin enables any third party to re-hash the same PDF document's and verify that they match, thereby confirming the files have not changed since the time of hashing. Even though the digital signatures already serve a similar purpose, the inclusion of SHA-256 hashes is another solid, "gold-standard" digital forensic measure to guarantee integrity, especially for the few files without digital signatures.

C | Chronological and Contextual Integrity

The dataset not only preserves individual files meticulously, but also the context of how and when they were collected. The '*05_MCRO_case-file-download-date-time.csv*' log captures the exact download date and time for each case's files, and delineates the three batch periods during which the scraping script ran. This means we have a timeline of collection: for instance, it documents that 11 cases were downloaded in the early hours of April 29, 2024, another batch of 33 cases on the afternoon of April 29, and the final 119 cases on April 30, 2024. Having this level of temporal detail is valuable in a forensic sense – it demonstrates that the data was gathered in a short, defined window (just before an announced system maintenance shutdown), and it has been static since. Moreover, by immediately producing an affidavit on May 3, 2024 that described this data, Guertin established a contemporaneous record. The short gap between data acquisition and affidavit filing (only about *3 days*) further solidifies chain-of-custody, as there was little opportunity for interference or modification in the interim. The dataset, the download logs, and the affidavit together paint a cohesive story of evidence handling: data pulled directly from the official source, quickly analyzed, and promptly entered into the court record as an affidavit. Any reviewer can follow this trail and see consistency at each step.

In conclusion, the file naming and metadata practices used in the MCRO dataset ensure that nothing is left to guesswork or assumption. Every file is self-descriptive and cross-documented, and all relevant metadata is preserved externally as well. This approach equips any subsequent investigator or court with the tools needed to verify the dataset's contents independently. From creation dates to cryptographic hashes, the provenance and integrity of each document can be confirmed without relying solely on trust. Such rigor is a hallmark of quality in digital forensic evidence handling.

VI. CONCLUSION

It is evident that the MCRO dataset is authentic, complete, and forensically sound. The collection's organization and verification measures go well beyond standard practice, reflecting Guertin's expert-level care in preserving digital evidence.

A | Key findings supporting the dataset's reliability

1. Comprehensive Coverage

All relevant case files (3,601 PDFs across 163 cases) have been captured and preserved, covering a broad timeframe and variety of document types. The dataset includes not just the individual filings but also contextual materials (case dockets and web pages) for each case, indicating no information was omitted. Even duplicate documents and edge cases are documented rather than swept aside.

2. Verified Authenticity

An overwhelming 99.6% of the PDFs retain original court digital signatures, confirming their authenticity and untouched state. The small fraction without signatures are transparently identified. In tandem with cryptographic hashes and recorded metadata, the dataset provides multiple layers of validation for each file's integrity.

3. Integrity of Organization

The exact folder/file structure of the evidence collection has been logged in detail, and file naming conventions have been preserved to carry inherent metadata (like timestamps and case IDs). This means the dataset can be navigated and understood with ease, and its structure can be compared to the original source layout for consistency. The inclusion of structured logs (CSV tables) for case indices, download times, and file counts by category further guarantees that the evidence set is internally consistent and well-documented.

4. Affidavit Corroboration

All statistical claims made in Guertin’s May 3, 2024 affidavit about this data are substantiated by the current dataset. The number of cases, distribution by year, and counts of documents by type are all exact matches. The only update was a corrected total file count (3,629 vs. 3,556) which has been duly noted and does not detract from the accuracy of the analysis. This high degree of correspondence validates the affidavit as a truthful summary of the data and enhances its evidentiary credibility.

B | A Highly Credible Foundation of Evidence

In summary, the MCRO dataset stands as a highly credible foundation of evidence for investigating the synthetic court records and judicial fraud that Guertin has uncovered. The dataset’s authenticity is beyond reasonable question – each file can be traced to its official origin and is backed by cryptographic proof. The careful cataloging and cross-checking performed ensure that any patterns or anomalies identified (such as unusual clusters of cases or document types) rest on a solid factual footing. From a digital forensic investigator’s perspective, the manner in which this dataset was collected, preserved, and audited is exemplary. It provides confidence that the evidence has not been tampered with and that it comprehensively represents the reality of the court records in question. Therefore, anyone evaluating the MCRO dataset can be assured of its integrity and usefulness as the cornerstone of Guertin’s case, and it should be afforded full weight in any legal or investigative proceedings that follow.

C | Source

MCRO Dataset CSV Tables

https://link.storjshare.io/s/jxhmrcnhsa2tdo5xfusewkoadnqq/evidence/MCRO/

https://link.storjshare.io/raw/jwmyznljwyyk4aqqhug2jp4filca/evidence/MCRO.zip

https://link.storjshare.io/s/ju3mf5uvdrmcbhch5ga3koduwp4q/evidence