DIGITAL FORENSIC ANALYSIS OF CASE DATASET INTEGRITY

I.   SUMMARY

This dataset contains structured information extracted from Minnesota Court Records Online (MCRO) criminal case dockets by Matthew Guertin using a custom Python/Selenium script. The extraction covers 163 Hennepin County criminal cases (2017–2023), with all available case filings downloaded as PDF (3,629 files, ~8,794 pages – 3,601 unique files after removing duplicates). Guertin parsed the saved HTML dockets into 12 interrelated CSV tables capturing case details, parties, events, and documents. The result is a comprehensive, cross-referenced docket dataset for forensic analysis, preserving original metadata (case numbers, timestamps, digital signatures, etc.) for authenticity verification.

II.   FILE INVENTORY

Each CSV file in 'CASE.zip' corresponds to a specific structured facet of the dockets. The files, record counts, and key fields are as follows:

1.   **'01_CASE_details.csv'**

163 records (one per case). Fields: Case number, status, assigned judge, origination date, case title, defendant name, location, etc..

2.   **'02_CASE_related.csv'**

101 records. Lists any related case numbers for each primary case (if applicable), with references to the same defendant.

3.   **'03_CASE_warrants.csv'**

675 records. Details of warrants issued in the cases (warrant IDs, issue/clear dates and times, status, issuing judge).

4.   **'04_CASE_listed-attorneys.csv'**

1,823 records. All attorneys of record for the cases, including defense and prosecution counsel (name, role, status, and whether listed as lead).

5. **'05_CASE_lead-attorneys.csv'**

326 records. The designated lead attorneys for each party in each case (defense and prosecution). Includes cases with no lead attorney noted (flagged accordingly).

6. **'06_CASE_charges.csv'**

266 records. All charges across the cases, with charge descriptions, Minnesota statute citations, and charge level (felony, misdemeanor, etc.).

7. **'07_CASE_interim-conditions.csv'**

3,679 records. Interim conditions imposed on defendants (e.g. conditional release terms), with the date set, judge ordering, condition description, and expiration date.

8. **'08_CASE_judicial-assignments.csv'**

292 records. History of judicial assignment for each case – including initial judge assignment dates and any reassignments (with dates and reasons).

9. **'09_CASE_docket-events.csv'**

11,841 records. All docket events from every case compiled chronologically, with each entry's date, description, presiding judicial officer, party (if relevant), and an indicator if a PDF document is associated. (Every event that had a downloadable file is mapped to its PDF, including filename and a Storj link for verification.)

10. **'10_CASE_hearings.csv'**

4,903 records. Scheduled and held hearings for all cases, with hearing dates, times, types (e.g. arraignment, competency hearing), locations, presiding officer, outcomes, and any ancillary actions.

11. **'11_CASE_clusters.csv'**

163 records. Identifies case clusters where the same defendant is involved in multiple cases. Each case entry notes if it's part of a cluster, and the total number of cases for that defendant (e.g. one defendant has 12 cases).

12. **'12_CASE_attorney-errors.csv'**

115 records. Logs of inconsistencies in attorney listings found in the dockets. Each row describes an anomaly such as an attorney appearing as both defense and

prosecution on the same case, being marked both active and inactive, or being flagged as both lead and non-lead counsel on a case.

III.   KEY FIELD STATISTICS

1.   **Total Cases**

163 unique case dockets are represented.

2.   **Unique Defendants**

85 distinct defendant names appear across the 163 cases (many individuals have multiple cases: only 40 cases involve a one-time defendant, while the other 123 case entries correspond to defendants with 2–12 cases each).

3.   **Unique Attorneys**

327 distinct attorney names are listed across all cases (including defense attorneys and prosecutors). This indicates an extensive cast of legal counsel involved in the docket data.

4.   **Unique Judicial Officers**

~80 different judges and referees are identified across the cases (via case assignments and event signatories), reflecting a broad range of court personnel appearing in the records.

5.   **Event Types**

103 distinct types of case events were identified in the docket events (e.g. various orders, notices, warrants, reports, motions, etc., as shown below).

IV.   EVENT TYPE FREQUENCY

The dataset captures a wide variety of docket filing types. The table below lists some of the most frequent event types recorded across all cases, along with their occurrence counts:

Event Type	Count of Entries
Hearing Held Remote	970
Notice of Remote Hearing with Instructions	694
Failure to Appear at a hearing	573
Hearing Held Using Remote Technology	519

Event Type	Count of Entries
Order – Evaluation for Competency to Proceed (Rule 20.01)	508
Bail to stand as previously ordered	468
Notice of Hearing	465
Hearing Held In-Person	436
Rule 20 Progress Report	405
Request for Continuance	398
Request for Interpreter	382
Found Incompetent	379
Order for Conditional Release	346
Warrant Issued	339
Rule 20 Evaluation Report	298

(The data includes many other filings with lower frequencies — e.g. "Warrant Cleared by Wt Office" (293 entries), generic "Motion" filings (274 entries), returned mail notices, orders appointing public defenders, etc. — totaling 103 distinct filing categories.)

The prevalence of Rule 20 competency proceedings (e.g. competency evaluations, progress reports, findings) and frequent failure-to-appear and warrant entries is notable from the counts above, indicating common themes across these synthetic cases.

V. TIMELINE OVERVIEW

The cases span a wide timeline. The earliest case in the dataset was filed on January 19, 2017, and the latest case was filed in late 2023 (November 14, 2023). Each year 2017 through 2023 is represented (e.g. 3 cases from 2017, 4 from 2018, … 39 from 2023). The docket activity for these cases runs from 2017 into 2024 – for example, some cases had hearings scheduled as far out as mid-2024 (the latest event date recorded is July 23, 2024, reflecting future hearings on the docket at the time of data capture). This timeline indicates the dataset covers approximately 7 years of case proceedings, from initial filings through ongoing court actions in 2024.

VI. DATA TRACEABILITY & INTEGRITY

This structured dataset is enriched with metadata to ensure evidentiary reliability of the extracted information:

A | Cross-Reference Links

Every record includes direct URL links (hosted via Storj) to the original MCRO content. For example, each case entry and event entry links back to the saved MCRO docket page or the specific PDF document for that filing. This means analysts can trace any data point directly to the source document or docket for verification. Additionally, case-level links to comprehensive zip files of all filings and HTML assets were maintained for each of the 163 cases.

B | Unique Identifiers

Key identifiers such as the official Case Number (e.g. 27-CR-XX-YYYY) are present in every table, ensuring that information across different tables can be joined and verified against the correct case. Each case also has a consistent sort index (zero-padded number form) to maintain sorting order. Docket events and related records carry indices and timestamps as in the original dockets, preserving the chronological order of occurrences.

C | Digital Signatures & Timestamps

The vast majority of the 3,629 downloaded PDF case files retain their original court digital signatures (99.6% had valid signatures). The embedded signing timestamps on those PDFs exactly match the timestamp suffixes in the files' names as downloaded. This provides strong cryptographic authentication that the documents are unaltered from their court-issued form and aligns with the recorded download time. Any files that did not retain a signature (only 16 out of 3,629) are explicitly identified in the separate signature report (in the MCRO dataset).

D | Quality and Completeness

The dataset captures every docket entry and file from the selected cases, enabling completeness checks. For instance, the 11,841 docket-event entries in the CSV exactly correspond to the 3,601 unique PDF filings downloaded (each file is mapped to its docket entry). This one-to-one mapping and the inclusion of both PDF page counts and file names for each event provide an audit trail to confirm that no documents are missing. Furthermore, any irregularities in the source data (such as the attorney listing contradictions) have been catalogued in the attorney-errors log for transparency.

VII. CONCLUSION

Overall, the structure and metadata richness of this dataset (unique IDs, timestamps, and source links for each entry) ensure that the extracted case docket content can be validated and cross-examined against the original court records with a high degree of confidence. The dataset's organization into thematic tables (cases, events, charges, etc.) offers a full picture of each case's timeline and participants, while the embedded references and signatures reinforce the trustworthiness of the data.

A | Source

CASE Dataset CSV Tables

https://link.storjshare.io/s/jxylovpvzqok36srek7ckcnuay6a/evidence/CASE/

https://link.storjshare.io/raw/jup3vkrw6mqnniigxlwa5qwye62q/evidence/CASE.zip

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