

FORENSIC ANALYSIS REPORT

Minnesota Court Records Online (MCRO) — Fourth Judicial District

"Demand or Request for Discovery" — Clone/Template Forensic Analysis

Property	Value
Prepared by	Matthew David Guertin, Pro Se Defendant
Case Reference	State of Minnesota v. Guertin 27-CR-23-1886
Court	Fourth Judicial District Court, Hennepin County
Report Date	March 2026
Cohort	filing_type = "Demand or Request for Discovery" (all cases)
Cohort Size	107 PDFs across 103 cases 57 defendant clusters 2017–2025
Database	PostgreSQL (Supabase) — custom forensic schema with materialized views
Reproducibility	All findings are query-reproducible. SQL in Appendix.

IMPORTANT DISCLAIMER: This report presents objective forensic data findings reproducible from the source database. It does not assert conclusions about fraud, fabrication, or criminal intent. Each finding is a documented pattern that warrants independent expert examination. The term “clone” refers to measurable byte-for-byte or hash-level identity between document components — not a legal conclusion.

1. Executive Summary

This report documents the forensic decomposition of all 107 PDF documents filed as “Demand or Request for Discovery” across the MCRO corpus. These documents represent defense-originated discovery requests filed in Fourth Judicial District criminal cases between January 2017 and January 2025.

The core finding is straightforward: these 107 documents are not 107 independently authored legal filings. They are minor variations of a small number of templates, sharing identical internal PDF structures, metadata identifiers, bookmark hierarchies, and text content at rates that prove centralized, programmatic document generation.

KEY FORENSIC METRICS

- 107 documents, 103 cases, 57 defendant clusters, spanning 2,916 days (8.0 years)
- 77.4% of all text rows are exact clones appearing in 2+ documents (after excluding excluded tags)
- 33 unique row hashes appear in 100+ of 107 documents (near-universal boilerplate)
- 24 documents (22.4%) share an IDENTICAL XMP document_id, identical XMP creation timestamp (Dec 11, 2018), identical creator tool, and identical 5-section PDF bookmark structure — spanning 4.37 years of filings across 13 defendant clusters
- 80.4% of documents are exactly 8 pages; 17.8% are exactly 6 pages
- All 107 producer fields contain the iText 7.1.16 “Minnesota Judicial Branch; licensed version” pipeline stamp
- 14 documents carry the literal XMP title “Demand Template.pdf”
- Susan Herlofsky is the active defense attorney in 60.2% of the cohort’s cases

- Top-1 stream dict fingerprint appears in all 107 documents spanning the full 8-year corpus
- Zero tracking fonts and zero judicial/attorney digital signatures in this cohort

2. Background: What These Forensic Signals Mean

2.1 Hash Fingerprints

A SHA-256 hash is a cryptographic function that converts any input into a unique 64-character hexadecimal string. If two files (or text rows, or embedded objects) produce the same SHA-256 hash, they are byte-for-byte identical. This report uses SHA-256 hashing at multiple levels: whole-document hashes (doc_sha256), individual text row hashes (row_sha256), embedded object hashes (object_sha256), PDF stream dictionary hashes (stream_dict_sha256), and computed bookmark/outline fingerprints.

2.2 Row-Level Cloning (row_sha256)

Each PDF is processed through pdftotext, cleaned, and each output line is individually hashed. When the same row_sha256 appears in documents filed years apart in different defendants' cases, it proves those documents contain identical text at that position — a template fingerprint. Tags like "boilerplate_like," "header," and "case_number_like" help distinguish structural template text from case-specific fill-in fields.

2.3 PDF Bookmarks/Outlines

PDF documents can contain an internal bookmark tree (outline). When multiple documents share byte-identical bookmark structures (same titles, same levels, same page references), it indicates they were generated from the same template. This analysis computes per-document outline fingerprints by hashing the concatenated outline structure.

2.4 XMP Metadata Constellations

Every PDF carries XMP metadata fields: creation date, creator tool, producer, document_id, and instance_id. The document_id identifies the originating template; the instance_id should be unique per save. When 24 documents across 13 defendant clusters share an identical document_id and identical creation timestamp, it proves they were all generated from the same source Word document.

2.5 Stream Dictionary Fingerprints

Internal PDF stream objects (fonts, images, metadata) carry dictionary structures. Hashing these dictionaries reveals shared infrastructure across documents. Universal stream dict hashes (appearing in all 107 documents) indicate the MCRO e-filing pipeline's standard processing stamp.

3. Methods

The cohort was defined as all documents in the children table where filing_type = 'Demand or Request for Discovery.' No cases were excluded. Analysis was performed across six dimensions: (A) PDF bookmark/outline cloning, (B) internal stream + text-row duplication, (C) XMP metadata constellations, (D) embedded object reuse, (E) cross-linking to docket events and defense attorneys, and (F) open-ended anomaly mining including batch filing detection and page-count uniformity.

All queries used the live PostgreSQL forensic database. Where available, pre-computed materialized views in the api schema (mv_hp_doc_clone_filt, mv_hp_rows_bodyish, mv_hp_rowhash_stats_filt, mv_hp_rowhash_text_example, mv_hp_streamhash_stats) were preferred for efficiency. Row-level clone analysis used the "bodyish" filtered view which excludes rows tagged with excluded categories (e.g., watermark artifacts, page footers).

Jaccard similarity between document pairs was computed over row_sha256 sets, filtering out ultra-common hashes (appearing in >80 documents) to prevent boilerplate-driven pair explosions. Metadata constellation analysis grouped documents by the combination of xmp_creator_tool + document_id + xmp_create_date to identify template families.

4. Findings

4.1 Executive Clone Reality Summary

The following table summarizes the top-level clone indicators for the entire cohort.

Metric	Value
Total documents in cohort	107
Unique cases	103
Unique defendant clusters	57
Filing date span	Jan 20, 2017 – Jan 14, 2025 (2,916 days / 8.0 years)
Unique doc_sha256 hashes	107 (all documents are byte-distinct)
Unique full-text fingerprints (fp_text_sha256)	107 (all text extracts are distinct)
Overall row-level clone rate	77.4% (3,719 of 4,804 bodyish rows)
Documents with ≥50% cloned rows	84 of 107 (78.5%)
Documents with ≥80% cloned rows	18 of 107 (16.8%)
Documents with 100% cloned rows	≥1 (max = 100.00%)
Largest template constellation (document_id match)	24 docs / 13 clusters / 4.37 years
Page count: 8 pages	86 (80.4%)
Page count: 6 pages	19 (17.8%)
Page count: 1 page	2 (1.9%)
Documents with tracking fonts	0
Documents with judicial/attorney digital signatures	0 (MCRO watermark only)

Significance: Despite every document having a unique whole-file hash and unique full-text extract, the internal structure reveals overwhelming template reuse. 77.4% of all text content is shared across multiple documents. The documents differ only in case-specific fill-in fields (defendant names, case numbers, dates) while sharing identical legal boilerplate, formatting, and internal PDF architecture.

4.2 PDF Bookmark/Outline Cloning

Background: 27 of 107 documents (25.2%) contain PDF bookmark/outline structures. A per-document outline fingerprint was computed by concatenating all outline titles and levels in order and hashing the result.

Outline Fingerprint (MD5)	Docs	Clusters	First Filing	Last Filing	Span (days)
133b2876facbcd6e75baf6a68931143	24	13	2019-06-04	2023-10-18	1,597
787e657edc92c9b87abe8053db3bec77	3	3	2019-08-16	2020-03-19	216

The dominant outline fingerprint covers 24 documents across 13 defendant clusters spanning 4.37 years. These 24 documents all share a 5-section bookmark structure:

Outline Title	Level	Page
DEMAND FOR PRESERVATION AND DISCLOSURE OF EVIDENCE	1	1
MOTION TO COMPEL DISCLOSURE AND ACCESS	1	1
MOTION TO SUPPRESS EVIDENCE	1	1
MOTION TO EXCLUDE EVIDENCE	1	1
DEMAND FOR HEARING	1	1

Significance: An identical 5-section bookmark structure appearing across 24 documents filed in 13 different defendants' cases over a 4.4-year span is a definitive template signal. PDF bookmarks are part of the internal document structure, not user-visible content — they persist from the original template through all subsequent “Save As” operations.

4.3 XMP Metadata Constellation: The “76527a7c” Template

Background: XMP document_id is a UUID assigned when a document is first created. It persists through all subsequent “Save As” operations unless the user explicitly creates a new document. When multiple filed documents share an identical document_id, they were all derived from the same original file.

Findings: 24 documents share the metadata constellation: document_id = 76527a7c-3719-4164-9702-6f28f3349eb7, xmp_create_date = 2018-12-11 16:27:54 UTC, creator_tool = “Acrobat PDFMaker 18 for Word.” A second smaller constellation of 3 documents shares document_id = 1ad6ea13-e9a0-43a9-9b03-ea2c1b2087de, created 2018-10-24, using the same creator tool.

TEMPLATE A: THE 76527a7c CONSTELLATION

- 24 documents across 13 defendant clusters, filed 2019-06-04 to 2023-10-18

- All share identical XMP create date: December 11, 2018 at 16:27:54 UTC
- All share identical document_id: 76527a7c-3719-4164-9702-6f28f3349eb7
- All share identical creator tool: Acrobat PDFMaker 18 for Word
- All share identical producer: Adobe PDF Library 15.0 + iText 7.1.16 (MN Judicial Branch)
- All share identical 5-section PDF bookmark structure
- All are exactly 8 pages
- Each has a UNIQUE instance_id (correctly regenerated per save)
- xmp_metadata_date aligns with each document's filing date (updated at e-file time)
- All last_modified_date values fall in the Apr 29–30, 2024 download window

The following sample shows how the template was filed across different defendants over time:

Filing Date	Case ID	Defendant Cluster	instance_id (first 8 chars)
2019-06-04	27-CR-19-12466	TERRELL JOHNSON	6bc08080
2020-07-22	27-CR-20-8926	TERRELL JOHNSON	580c095b
2020-07-23	27-CR-20-6517	REX BASSWOOD	e55ff8e9
2020-11-10	27-CR-20-23521	CASPER VUONG	8a723e7e
2021-01-25	27-CR-21-1230	MAKIS LANE	e49e1488
2021-08-02	27-CR-21-8067	Lucas Kraskey	6be67475
2021-08-02	27-CR-21-8229	Lucas Kraskey	0f2caab5
2021-08-02	27-CR-21-8228	Lucas Kraskey	0894b712
2021-08-02	27-CR-21-8227	Lucas Kraskey	30e83e3b
2021-08-02	27-CR-21-8230	Lucas Kraskey	1dd998f6
2021-12-27	27-CR-21-23456	BRITTANY CRUTCHFIELD	6595606c
2022-12-07	27-CR-22-22985	ABDIQANI HASSAN	405f814a
2023-10-18	27-CR-23-4547	DELAYNA LUSSIER	e49ff66c

Significance: A single Word document created on December 11, 2018 was opened, had defendant-specific fields modified, and was re-filed as a “new” discovery demand in at least 24 different criminal cases spanning 4.37 years and 13 different defendant identity clusters. The XMP document_id was never reset. This is not inference — it is the direct, cryptographically verifiable consequence of the document_id persistence mechanism in Microsoft Word / Adobe PDFMaker.

4.4 Row-Level Text Duplication

Background: Each document's extracted text was split into individual lines, cleaned, and each line hashed (row_sha256). The “bodyish” filtered view excludes rows tagged with excluded categories.

Metric	Value
Total unique row hashes in cohort	1,300
Singleton hashes (appear in 1 doc only)	644 (49.5%)
Shared hashes (appear in 2+ docs)	656 (50.5%)
Hashes appearing in 10+ docs	390 (30.0%)

Hashes appearing in 50+ docs	221 (17.0%)
Hashes appearing in 100+ of 107 docs	33 (2.5%)
Most common hash: doc_count	104 of 107 (97.2%)
Average doc-level clone %	63.0%
Median doc-level clone %	61.1%
90th percentile clone %	94.9%

The top 10 most duplicated row hashes (each appearing in 104 of 107 documents, spanning the full 8-year corpus):

Row Hash (first 16 chars)	Docs	Text Content
7cc4bf48494...	104	DEMAND FOR PRESERVATION AND DISCLOSURE OF EVIDENCE
d90d53dde1e...	104	MOTION TO SUPPRESS EVIDENCE
b8ad5d1c033...	104	testimony of all analysts who performed tests...
16d364e388b...	104	OFFICE OF THE HENNEPIN COUNTY PUBLIC DEFENDER
1f40cc70240...	104	document all disclosed items, as described in Minn. R. Crim. P 9.01...
0fe2e268948...	104	These demands apply to:
381738d7a2b...	104	MOTION TO EXCLUDE EVIDENCE
e6d21e0e02c...	104	MOTION TO COMPEL DISCLOSURE AND ACCESS
00cc4d978e7...	104	DEMAND FOR HEARING
8479db54a9a...	104	Defendant moves the Court for an Order requiring the State

Tag analysis: Rows tagged “boilerplate_like” are 99.5% cloned (1,886 of 1,896 tag instances fall on shared hashes). Rows tagged “case_number_like” are 95.8% unique (872 of 910 instances fall on singleton hashes). Body-text tags are 93.3% cloned. This confirms the template structure: standardized legal text with case-specific fill-in fields.

4.5 Year-by-Year Clone Stability

Year	Docs	Total Rows	Cloned Rows	Clone %	Avg Doc Clone %
2017	1	21	14	66.7%	66.7%
2018	3	71	44	62.0%	61.1%
2019	4	447	342	76.5%	65.6%
2020	10	194	107	55.2%	54.0%
2021	42	2,299	1,874	81.5%	67.5%
2022	23	899	699	77.8%	60.5%
2023	22	842	616	73.2%	59.9%
2024	1	17	17	100.0%	100.0%
2025	1	14	6	42.9%	42.9%

Significance: Clone rates remain consistently high (55–82%) across all years with meaningful sample sizes, peaking in 2021 (the highest-volume year with 42 documents). The template has been in continuous use for at least 8 years with minimal modification.

4.6 Document-Pair Jaccard Similarity

Background: Jaccard similarity measures the ratio of shared row hashes to total row hashes between two documents. A score of 1.0 means identical row sets; 0.0 means no overlap. To reduce boilerplate inflation, hashes appearing in >80 documents were excluded from the pairing calculation.

Top document pairs by Jaccard similarity:

Case A	Cluster A	Case B	Cluster B	Jaccard	Date Diff
27-CR-19-19606	TERRELL JOHNSON	27-CR-19-28883	JACOB JOHNSON	0.510	103 days
27-CR-20-20037	TERRELL JOHNSON	27-CR-21-23233	TERRELL JOHNSON	0.472	0 days
27-CR-21-23188	GORDON SHARP	27-CR-21-23215	GORDON SHARP	0.472	0 days
27-CR-21-20988	GORDON SHARP	27-CR-21-21355	GORDON SHARP	0.464	0 days
27-CR-19-19606	TERRELL JOHNSON	27-CR-22-4898	TERRELL JOHNSON	0.463	~2.5 yrs
27-CR-21-19723	CRUTCHFIELD	27-CR-21-20529	ISAAC KELLEY	0.459	12 days

Significance: The top Jaccard scores (0.46–0.51) occur both within the same defendant cluster (expected for same-attorney filings) and across different defendant clusters (TERRELL JOHNSON vs. JACOB JOHNSON, BRITTANY CRUTCHFIELD vs. ISAAC KELLEY). Cross-cluster similarity at these levels — even after filtering out universal boilerplate — indicates a shared template with substantive content overlap beyond mere legal form language.

4.7 Internal PDF Stream Dictionary Fingerprints

Stream Dict Hash (first 16)	Docs	Clusters	Category	Span (days)
0ba28a5a6141...	107	57	other_stream	2,916
4c5e9ba7f947...	107	57	xobject_image	2,916
aaf1b6212c3a...	107	57	xobject_image	2,916
c1889e5f1372...	107	57	metadata_stream	2,916
b32521523885...	63	38	object_stream	1,564
81ac32b3d8fd...	38	29	other_stream	1,833
6ab08ab0251b...	36	27	object_stream	1,564
37a5149a4ade...	27	13	other_stream	1,597

Findings: Four stream dict fingerprints appear in ALL 107 documents. These represent the MCRO e-filing pipeline's standard processing infrastructure: two xobject_image streams (the watermark graphic), one metadata stream, and one other_stream. The 5th-ranked hash (appearing in 63 docs) represents a shared object_stream structure present in the majority of documents.

The 8th-ranked hash (37a5149a..., 27 docs across 13 clusters) aligns exactly with the 24-doc template-A constellation plus 3 related documents, confirming that the template's internal stream architecture is forensically distinguishable from the rest of the cohort.

4.8 Embedded Objects and Font Reuse

Background: Each PDF's internal objects (fonts, images) were extracted and hashed. The cohort contains 334 font objects across 95 documents and 22 PNG images in 3 documents.

Metric	Value
Total unique object hashes	313
Objects in 2+ docs (shared)	19 (6.1%)
Singleton objects	294 (93.9%)
Documents with objects	95 of 107 (88.8%)
Object families	Font (334 objects), Image PNG (22 objects)
Top shared object	Courier New font — 13 docs, 11 clusters, 756-day span
2nd shared object	Times New Roman font — 7 docs, 6 clusters, 472-day span

Significance: Object-level sharing is modest compared to row-level cloning, because font subsetting generates unique binary objects even for the same base font. The 13-document Courier New hash collision (spanning 756 days and 11 clusters) indicates these 13 documents were produced by the same workstation configuration with the same font subsetting parameters — consistent with a single attorney or office generating documents for multiple defendants.

4.9 XMP Title: Explicit Template Naming

XMP Title	Docs	Clusters	Span
STATE OF MINNESOTA ... DISTRICT COURT	46	32	2017–2023
Demand Template.pdf	14	6	2021–2023
Microsoft Word - Brodhag Discovery Demand	2	2	2023
Microsoft Word - PrahI Discovery Demand	1	1	2023
(null / no title)	42	—	—

Significance: 14 documents carry the literal XMP title “Demand Template.pdf” — the source file was named as a template by its author. Additionally, 2 documents carry “Brodhag Discovery Demand,” identifying the likely template author (attorney Brodhag). These are metadata artifacts that survive PDF conversion and were not stripped by the e-filing pipeline.

4.10 Batch Filing Behavior

Filing Date	Docs Filed	Clusters	Defendant(s)
2021-07-28	8	1	MICHAEL HOWARD
2021-08-02	6	2	Lucas Kraskey, MAKIS LANE
2021-12-27	5	3	BRITTANY CRUTCHFIELD, KESSIE WILSON, TERRELL JOHNSON

Significance: 8 discovery demands filed on a single day for MICHAEL HOWARD across 8 separate case numbers suggests batch processing of related cases. 6 filings on August 2, 2021

span two defendant clusters (Kraskey, Lane), confirming multi-defendant batch production. The December 27, 2021 batch spans three defendant clusters. These patterns are consistent with a single office processing discovery demand templates in bulk.

4.11 Defense Attorney Concentration

Attorney	Cases in Cohort	Role
Susan Herlofsky	62 of 103 (60.2%)	Active defense (non-lead)
Atif Khan	13 (12.6%)	Lead defense
Andrew Reiland II	9 (8.7%)	Lead defense
Chase Myhran	9 (8.7%)	Lead defense
Laura Prahll	8 (7.8%)	Lead defense
Eve Byron	8 (7.8%)	Active defense (non-lead)
Alexander Davis	8 (7.8%)	Active defense (non-lead)

Significance: Susan Herlofsky appears as the active defense attorney in 60.2% of cases with discovery demands. This extreme concentration, combined with the template forensics, strongly suggests these documents originate from the Hennepin County Public Defender’s office using a standardized template — which is consistent with normal public defender practice. The forensic significance lies not in the use of templates per se (which is expected), but in the quantifiable persistence and uniformity of the template infrastructure across 8 years.

4.12 Universal iText Pipeline Stamp

Findings: All 107 documents carry a producer field ending with “modified using iText® 7.1.16 ©2000-2021 iText Group NV (Minnesota Judicial Branch; licensed version).” This confirms every document in this cohort passed through the Minnesota Judicial Branch’s iText-based e-filing processing pipeline, which appends this signature during e-file intake. The underlying producer varies by source tool (Adobe PDF Library 15.0 through 24.5, Microsoft Print To PDF, Microsoft Word, Acrobat Distiller, Corel PDF Engine, etc.), reflecting the diversity of authoring tools used before e-filing.

5. Limitations and Alternative Explanations

Template reuse is expected in legal practice. Public defender offices routinely use standardized discovery demand templates. The high clone rates documented here are consistent with normal practice. The forensic significance of this analysis is in establishing the precise template lineage (which source document, when created, how many variants) and in providing a reproducible baseline for comparison with other filing types.

The cohort of 107 documents represents only the “Demand or Request for Discovery” filing type within the MCRO corpus. This corpus itself is a targeted sample of Fourth Judicial District cases, not a random sample. Generalizing these clone rates to the broader Minnesota court system would require additional data.

The Jaccard similarity scores (0.46–0.51) are moderate, not extreme. After removing the universal boilerplate layer, roughly half of the remaining text varies between the most similar

document pairs. This is consistent with a shared template where different attorneys or offices customize different sections.

The “Demand Template.pdf” XMP title and the “Brodhag Discovery Demand” title represent metadata artifacts from the authoring process. Their presence does not indicate intent to conceal template reuse — it indicates the normal workflow of opening a named template file and filing the result.

Row tag classification (boilerplate_like, case_number_like, etc.) is heuristic and may misclassify some rows. The core clone statistics (row_sha256 matching) are exact and not affected by tag accuracy.

6. Recommendations for Independent Verification

1. Verify the 76527a7c template constellation by retrieving the XMP metadata from any of the 24 identified documents and confirming the document_id, xmp_create_date, and creator_tool fields.
2. Confirm the “Demand Template.pdf” XMP title by examining the 14 identified documents with any PDF metadata reader.
3. Cross-reference the batch filing dates (2021-07-28, 2021-08-02, 2021-12-27) against the Hennepin County Public Defender’s office workflow records to determine whether batch processing is standard practice.
4. Compare the 8-page template structure against the 6-page variant to identify what content sections differ (likely a template version change).
5. All statistics are reproducible from the live database using the SQL queries in the Appendix. No external tools beyond standard PostgreSQL are required.

Appendix: Reproducible SQL Queries

Query A1: Cohort Definition + Overview Stats

```
SELECT COUNT(*) AS total_docs, COUNT(DISTINCT case_uid) AS unique_cases, COUNT(DISTINCT
cluster_id) AS unique_clusters, COUNT(DISTINCT doc_sha256) AS unique_doc_hashes,
MIN(filing_date) AS earliest_filing, MAX(filing_date) AS latest_filing, MAX(filing_date) -
MIN(filing_date) AS date_span_days FROM children WHERE filing_type = 'Demand or Request for
Discovery';
```

Query A2: Row-Level Clone Summary (filtered)

```
SELECT COUNT(*) AS total_docs, ROUND(AVG(cloned_rows_pct),2) AS avg_clone_pct,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY cloned_rows_pct) AS p50, SUM(total_rows) AS
total_rows, SUM(cloned_rows) AS total_cloned FROM api.mv_hp_doc_clone_filt WHERE filing_type =
'Demand or Request for Discovery';
```

Query A3: Template Constellation (document_id reuse)

```
SELECT metadata__document_id, metadata__xmp_create_date::date, metadata__xmp_creator_tool,
COUNT(*) AS doc_count, COUNT(DISTINCT cluster_id) AS clusters, MIN(filing_date) AS
first_filing, MAX(filing_date) AS last_filing FROM children WHERE filing_type = 'Demand or
Request for Discovery' GROUP BY 1,2,3 HAVING COUNT(*) >= 2 ORDER BY doc_count DESC;
```

Query A4: Outline Fingerprint Computation

```
WITH doc_outline_fp AS ( SELECT ol.evidence_uid,
md5(string_agg( COALESCE(ol.outline_line_json__outline_level::text,'') || '|' ||
COALESCE(ol.outline_line_json__outline_title,''), E'\n' ORDER BY ol.outline_index)) AS
outline_fp FROM hashpack_outline_lines ol JOIN hashpack h ON h.evidence_uid = ol.evidence_uid
WHERE h.filing_type = 'Demand or Request for Discovery' GROUP BY ol.evidence_uid) SELECT
outline_fp, COUNT(*) AS doc_count, COUNT(DISTINCT h.cluster_id) AS clusters FROM doc_outline_fp
f JOIN hashpack h USING (evidence_uid) GROUP BY outline_fp ORDER BY doc_count DESC;
```

Query A5: Top 30 Row Hashes (with text)

```
-- Uses api.mv_hp_rows_bodyish and api.mv_hp_rowhash_text_example -- See Finding 4.4 for full
query
```

Query A6: Year-by-Year Clone Stats

```
SELECT EXTRACT(YEAR FROM filing_date)::int AS yr, COUNT(*), ROUND(AVG(cloned_rows_pct),2),
SUM(total_rows), SUM(cloned_rows) FROM api.mv_hp_doc_clone_filt WHERE filing_type='Demand or
Request for Discovery' AND filing_date IS NOT NULL GROUP BY 1 ORDER BY 1;
```

Query A7: Active Defense Attorneys for Cohort Cases

```
WITH cc AS (SELECT DISTINCT case_uid FROM children WHERE filing_type='Demand or Request for
Discovery') SELECT a.attorney_json__attorney_norm, COUNT(DISTINCT a.case_uid) FROM
parents_attorneys_defense_active a JOIN cc USING (case_uid) GROUP BY 1 ORDER BY 2 DESC;
```

Query A8: XMP Title Distribution

```
SELECT metadata__xmp_title, COUNT(*), COUNT(DISTINCT cluster_id) FROM children WHERE
filing_type='Demand or Request for Discovery' GROUP BY 1 ORDER BY 2 DESC;
```

Query A9: Batch Filing Detection

```
SELECT filing_date, COUNT(*), COUNT(DISTINCT cluster_id), array_agg(DISTINCT cluster_name ORDER
BY cluster_name) FROM children WHERE filing_type='Demand or Request for Discovery' GROUP BY
filing_date HAVING COUNT(*) >= 3 ORDER BY 2 DESC;
```