MCRO | Doc-Set Group Forensics

Part 2 of 5: Hashpack Text, Stream & Outline Hash Forensics

8 Doc-Set Groups · 1,043 Unique Documents · Content Homogeneity & Cross-Group Sharing Analysis

Field	Value
Report Date	March 5, 2026
Series	Part 2 of 5 — Content Hash Forensics (Parts 1–5 inherit cohort from Part 1)
Data Source	Minnesota Court Records Online (MCRO) — Public Records
Database	PostgreSQL (Supabase, ibfmjtwahkwqzcmeyqii, us-west-2) — Live query
Corpus Scope	8 doc_set-tagged cohorts; Anchor CTE identical to Part 1
Total Unique Docs	1,043 (cohort count verified against Part 1 before analysis)
Primary Tables	hashpack, hashpack_rows, hashpack_page_content_streams, hashpack_outline_lines
Reproducibility	All statistics derived from live DB queries. SQL in Appendix.
Prepared by	Matthew David Guertin, Pro Se — Case 27-CR-23-1886

IMPORTANT DISCLAIMER

This report presents objective forensic data findings reproducible from a live PostgreSQL database. It does not assert conclusions about guilt, fabrication of records, fraud, or criminal intent by any party.

This is Part 2 of a 5-part forensic series. Part 1 established the definitive cohort census for 8 doc_set groups. All statistics in this report were computed fresh from the live MCRO forensic database immediately before report generation. No numbers from Part 1 or any prior analysis have been carried forward as facts — every figure herein is freshly queried.

CONTINUITY NOTE: Cohort counts were verified against Part 1 before any analysis proceeded. All 8 group sizes match exactly. The Anchor CTE uses column name 'value' (not 'doc_set') as discovered in Part 1.

Findings represent measurable patterns in document hash content. Each finding requires independent expert examination before use in formal legal proceedings.

EXECUTIVE SUMMARY

- Cohort Continuity Verified: All 8 group counts match Part 1 exactly (1,043 unique docs, 1,091 per-group assignments). Hashpack coverage is 100% across every group — no gaps in the hash evidence chain.
- AUTH_Judith-Cole: Extreme Exact-Clone Rate. 73.2% of its 123 documents share their fp_text_sha256 with at least one other document in the group — the highest clone rate by a wide margin. The largest clone group contains 20 identical-text documents. The median document in this group has 0.0% of its text rows unique to it within the group.
- AUTH Groups Dominate Template Reuse. The four AUTH groups fill the top-4 slots for within-group row-level template reuse: AUTH_Judith-Cole (91.2%), AUTH_Amanda-Burg (91.1%), AUTH_Megan-Larison-DHS (82.1%), AUTH_Alisha-Nehring (67.6%). All four produce documents that are overwhelmingly composed of repeated, shared text.
- 7 Row Hashes Are Universal — Present in All 8 Groups. The text rows 'State of Minnesota', 'Filed in District Court', 'STATE OF MINNESOTA DISTRICT COURT', 'Judge of District Court', '2', 'ORDER', and 'Defendant.'

appear in every one of the 8 groups and span the full 8.5-year corpus window (3,104 days). The most universal hash — 'State of Minnesota' — appears in 641 documents.

- Strongest Cross-Group Pairing: AUTH_Alisha-Nehring ↔ META_ESolutions-Signature (Jaccard 18.63%). These two groups share 960 distinct row hashes — the highest cross-group textual overlap. The second-strongest pair is META_ESolutions-Signature ↔ META_Signed-after-eFile (17.66%, 802 shared hashes). Both META groups are also structurally coupled.
- AUTH_Amanda-Burg ↔ AUTH_Megan-Larison-DHS: Unexpected AUTH–AUTH Coupling (Jaccard 16.74%). Despite being distinct author groups, these two share 81 row hashes with 16.74% Jaccard similarity — higher than most META–META pairs. This indicates shared template text between two nominally independent document sets.
- META_MSIP: Distinct Stream Pipeline. The dominant page content stream hash for META_MSIP appears in 26 of 34 documents (76.5%) — the highest stream-hash concentration of any group — and is not shared with any other group. META_MSIP also has the highest outline coverage (82.4%), with exclusively appellate-format bookmark titles ('O R D E R', 'STATE OF MINNESOTA IN COURT OF APPEALS').
- META_Signed-after-eFile & META_MSIP: Highest Document-Level Uniqueness. These two groups have the lowest clone rates (0.0%) and highest per-document row uniqueness scores (63.8% and 58.8% mean), making them the most content-diverse groups in the corpus — contrasting sharply with the three AUTH groups at the bottom.
- A Row Hash Encoding 'Defendant is presently incompetent to stand trial.' appears in 51 documents across 4 groups spanning 2,990 days (8.2 years) — an 8-year boilerplate template line tied to the core Rule 20 pipeline language.

TABLE ES-1: Cohort Verification & Key Hash Metrics Per Group

Group	Docs	Hashpack Coverage	Distinct fp_text_sha256	Exact-Clone Rate %	Template Reuse %	Mean Doc Uniqueness %	Outline Coverage %
AUTH_Alisha-Nehring	29	100%	23	31.0%	67.6%	17.5%	27.6%
AUTH_Amanda-Burg	75	100%	70	10.7%	91.1%	3.7%	0.0%
AUTH_Judith-Cole	123	100%	50	73.2%	91.2%	5.6%	22.0%
AUTH_Megan-Larison-DHS	20	100%	20	0.0%	82.1%	7.7%	0.0%
META_Created-after-eFile	656	100%	521	26.7%	63.1%	21.7%	2.0%
META_ESolutions-Signature	134	100%	110	27.6%	58.5%	29.1%	12.7%
META_MSIP	34	100%	34	0.0%	30.2%	58.8%	82.4%
META_Signed-after-eFile	20	100%	20	0.0%	25.5%	63.8%	20.0%

⚑ KEY METRICS — NUMBERS THAT DEMAND ATTENTION

- AUTH_Judith-Cole exact-clone rate: 73.2% — nearly 3 in 4 documents are byte-for-byte identical in text content to at least one other document in the group.
- AUTH_Judith-Cole median per-document uniqueness: 0.0% — the typical Cole document contains ZERO text rows not also present in other Cole documents.
- 7 row hashes appear in ALL 8 groups spanning 3,104 days (8.5 years). These are structural boilerplate invariants of the entire pipeline.
- AUTH_Alisha-Nehring ↔ META_ESolutions-Signature Jaccard: 18.63% — 960 shared text rows between a named-author group and a metadata-signature group.
- META_Created-after-eFile top stream hash: present in 298 of 656 documents (45.4% of the group).
- META_MSIP dominant stream hash: present in 26 of 34 documents (76.5%) — exclusively within MSIP, indicating a distinct rendering pipeline.
- The row 'Defendant is presently incompetent to stand trial.' appears in 51 documents across 4 groups spanning 8.2 years.
- AUTH_Megan-Larison-DHS stream layer: 41 page-content streams across 20 documents — every stream hash is unique (zero reuse between docs).

METHODS

Cohort Definition & Anchor CTE

This report uses the identical Anchor CTE established in Part 1. The cohort is defined by joining children_doc_set (column: value) to children on evidence_uid, filtering to the 8 named doc_set groups. Before any analysis, cohort counts were verified against Part 1 using COUNT(DISTINCT evidence_uid) per group. All 8 counts matched exactly. The Anchor CTE is reproduced in full in Appendix A1.

Multi-Group Membership

As established in Part 1, 48 documents belong to more than one doc_set group. All per-group statistics in this report count a document in every group it belongs to (full membership, additive). Where multi-group membership could skew results, both the additive and deduplicated perspectives are noted.

Schema Discovery

Before executing any analytics, column names were verified via information_schema scouts. Key discoveries: (1) hashpack rows use column text_clean (not row_text as in the schema reference); (2) hashpack uses n_pages (not page_count) and fp_streams_other_stream_count for stream metrics; (3) hashpack_outline_lines does not contain a hash column — bookmark deduplication uses outline_line_json__outline_title (exact text); (4) the api schema contains pre-built views including hashpack_doc_clone_raw, hashpack_rowhash_stats_raw, hashpack_streamhash_stats, and hashpack_outline_stats — these were examined for semantic consistency before use.

Analyses Performed

Six analyses were executed: (1) Hashpack coverage and fp_text_sha256 exact-clone detection — counting documents sharing identical whole-document text fingerprints within each group; (2) Row-level hash collision rates — within-group template reuse ratios, tier distributions (unique / 2–5 / 6–20 / 20+ documents), top boilerplate rows per group, and per-document uniqueness scores; (3) Cross-group Jaccard similarity matrix — computing intersection and union sizes of distinct row_sha256 sets for all 28 group pairs; (4) Stream dictionary fingerprints — using hashpack_page_content_streams to profile rendering pipeline signatures per group; (5) Outline/bookmark hash analysis — coverage and title frequency using hashpack_outline_lines; (6) Content Uniqueness Cards — synthesizing all metrics into per-group profiles.

Tables and Views Used

Primary tables: children_doc_set (cohort definition, column: value), children (document metadata), hashpack (fp_text_sha256, fp_streams_other_stream_count, n_pages), hashpack_rows (row_sha256, text_clean), hashpack_page_content_streams (decoded_sha256, page_num), hashpack_outline_lines (outline_line_json__outline_title). The api schema views were inspected for column structure but direct table queries were used for reproducibility. No children_objects, children_signatures_full_report, or children_doc_strings tables were used (reserved for Parts 3–5).

FINDINGS

Finding 1: Hashpack Coverage & Exact-Clone Detection (fp_text_sha256)

Background

The fp_text_sha256 field in the hashpack table stores a SHA256 fingerprint of each document's full extracted text content (computed from the normalized text stream). Documents sharing the same fp_text_sha256 are byte-for-byte identical in extractable text — exact content clones. This analysis measures, for each group, what fraction of documents are clones of at least one other document in the same group.

Findings

Hashpack coverage is 100% across all 8 groups. Every document has an fp_text_sha256 fingerprint. The exact-clone landscape is highly polarized:

Group	Total Docs	Distinct fp_text_sha256	Clone Groups (≥2 docs)	Largest Clone Group	Exact-Clone Rate %
AUTH_Judith-Cole	123	50	17	20 docs	73.2%
AUTH_Alisha-Nehring	29	23	3	4 docs	31.0%
META_ESolutions-Signature	134	110	13	5 docs	27.6%
META_Created-after-eFile	656	521	40	15 docs	26.7%
AUTH_Amanda-Burg	75	70	3	3 docs	10.7%
AUTH_Megan-Larison-DHS	20	20	0	—	0.0%
META_MSIP	34	34	0	—	0.0%
META_Signed-after-eFile	20	20	0	—	0.0%

AUTH_Judith-Cole stands out: 73 of its 123 documents (59.3%) share the same fp_text_sha256 as at least one other document within the group when counting only clone-group members. Restated: only 33 distinct text fingerprints exist among 123 documents (50 unique, but 17 hashes represent clone groups). The largest clone group of 20 documents means 20 separate court filings contain byte-for-byte identical text content. META_Created-after-eFile, despite being 5× larger, has only 40 clone groups with a maximum clone size of 15 — a lower concentration.

Significance

Within AUTH_Judith-Cole, 73.2% of documents share text fingerprints with at least one other document in the group. For a set of 123 documents representing distinct court orders filed across different cases and dates, this rate of textual identity indicates systematic template-driven production. Three groups (AUTH_Megan-Larison-DHS, META_MSIP, META_Signed-after-eFile) show zero within-group clones — each document in these groups has a unique text fingerprint.

Connecting Context

Part 1 found that AUTH_Judith-Cole has 85.4% filing-type concentration in 'Findings and Order' — a single document type. This Finding explains the mechanism: those near-identical orders are being generated from a small number of master templates. The 17 clone groups map to functional template variants within that filing type.

Finding 2: Within-Group Row-Level Template Reuse Rates

Background

Where fp_text_sha256 equality identifies documents identical at the whole-document level, row-level SHA256 analysis detects shared text at finer granularity. Each row of extracted text has its own SHA256 hash (row_sha256). When the

same row_sha256 appears across multiple documents in a group, those documents share that exact text line — evidence of template reuse even when documents are not identical overall.

Findings

Template Reuse Ratios (within-group):

Group	Total Rows	Distinct Row SHA256s	Within-Group Uniqueness %	Template Reuse %	Mean Doc Uniqueness %	Median Doc Uniqueness %
AUTH_Judith-Cole	20,067	1,773	8.8%	91.2%	5.6%	0.0%
AUTH_Amanda-Burg	4,370	388	8.9%	91.1%	3.7%	3.0%
AUTH_Megan-Larison-DHS	988	177	17.9%	82.1%	7.7%	3.6%
AUTH_Alisha-Nehring	5,028	1,628	32.4%	67.6%	17.5%	6.6%
META_Created-after-eFile	37,840	13,947	36.9%	63.1%	21.7%	12.9%
META_ESolutions-Signature	10,803	4,485	41.5%	58.5%	29.1%	20.0%
META_MSIP	3,380	2,360	69.8%	30.2%	58.8%	59.9%
META_Signed-after-eFile	1,152	858	74.5%	25.5%	63.8%	71.8%

Row Reuse Frequency Tier Distribution (distinct row hashes per tier, by group):

Group	Unique (1 doc)	2–5 docs	6–20 docs	20+ docs (Boilerplate)
AUTH_Judith-Cole	889	438	261	185
AUTH_Amanda-Burg	102	127	96	63
AUTH_Megan-Larison-DHS	60	48	69	0
AUTH_Alisha-Nehring	397	1,146	78	7
META_Created-after-eFile	10,421	2,476	834	216
META_ESolutions-Signature	2,693	1,557	202	33
META_MSIP	2,148	174	26	12
META_Signed-after-eFile	763	84	11	0

AUTH_Judith-Cole has 185 distinct row hashes appearing in 20 or more documents — the highest boilerplate count of any group. AUTH_Megan-Larison-DHS and META_Signed-after-eFile have zero 20+ boilerplate rows, indicating shallower template reuse. The top boilerplate rows per group are legal form text: jurisdictional headers, party labels, and structural markers.

Top Boilerplate Rows per Group (top-1 by doc count):

Group	Top Row Text	Doc Count
AUTH_Alisha-Nehring	'Plaintiff,'	29 / 29
AUTH_Amanda-Burg	'2'	70 / 75
AUTH_Judith-Cole	'3'	123 / 123
AUTH_Megan-Larison-DHS	'2. By presentation of a copy of this order...'	13 / 20
META_Created-after-eFile	'State of Minnesota'	350 / 656
META_ESolutions-Signature	'State of Minnesota'	131 / 134
META_MSIP	'2'	30 / 34
META_Signed-after-eFile	'Filed in District Court'	18 / 20

Significance

The AUTH groups exhibit the lowest per-document uniqueness scores (mean 3.7–17.5%), with AUTH_Amanda-Burg and AUTH_Judith-Cole having median per-document uniqueness below 4%. A median of 0.0% for AUTH_Judith-Cole means that in more than half the documents in this group, every single text row appears in at least one other document in the group — leaving nothing that is textually unique to that individual document. AUTH_Megan-Larison-DHS's top-1 row samples complete legal clause text ('By presentation of a copy of this order...') indicating a specialized DHS-specific template with full paragraph-level boilerplate.

Connecting Context

Part 1 identified AUTH_Amanda-Burg as spanning only a 724-day window (2021–2023) — a compressed temporal footprint. This Finding confirms that within that window, documents are almost entirely composed of shared template text (91.1% reuse rate). AUTH_Judith-Cole's 2,128-day span (Part 1) combined with 91.2% row reuse indicates a stable, long-running template maintained across 5+ years.

Finding 3: Cross-Group Row Hash Sharing Matrix

Background

Jaccard similarity between two groups is computed as: |intersection of distinct row_sha256 sets| / |union of distinct row_sha256 sets|. A value of 0% means no shared text rows whatsoever; 100% would mean textually identical groups. Because the groups represent different document categories and authors, high Jaccard values indicate shared production infrastructure, template lineage, or common document origin.

Findings

8×8 Jaccard Similarity Matrix (%) — Row Hash Intersection / Union:

Group	AN	AB	JC	ML	CE	ES	MS	SF
Alisha-Nehring	—	1.41	1.07	0.95	1.02	18.63	1.09	7.53
Amanda-Burg	1.41	—	0.51	16.74	0.17	3.07	0.44	5.33
Judith-Cole	1.07	0.51	—	0.52	1.73	2.25	0.41	3.14
Megan-Larison-DHS	0.95	16.74	0.52	—	0.12	1.06	0.40	4.23
Created-after-eFile	1.02	0.17	1.73	0.12	—	1.86	0.45	1.86
ESolutions-Signature	18.63	3.07	2.25	1.06	1.86	—	2.87	17.66
MSIP	1.09	0.44	0.41	0.40	0.45	2.87	—	0.91
Signed-after-eFile	7.53	5.33	3.14	4.23	1.86	17.66	0.91	—

Key: AN=AUTH_Alisha-Nehring, AB=AUTH_Amanda-Burg, JC=AUTH_Judith-Cole, ML=AUTH_Megan-Larison-DHS, CE=META_Created-after-eFile, ES=META_ESolutions-Signature, MS=META_MSIP, SF=META_Signed-after-eFile. Color: red ≥15%, orange 10–15%, yellow 5–10%, pale 1–5%, white <1%.

Ranked Cross-Group Pairs by Jaccard Similarity:

Rank	Group A	Group B	Intersection (shared row hashes)	Jaccard %
1	AUTH_Alisha-Nehring	META_ESolutions-Signature	960	18.63%
2	META_ESolutions-Signature	META_Signed-after-eFile	802	17.66%
3	AUTH_Amanda-Burg	AUTH_Megan-Larison-DHS	81	16.74%
4	AUTH_Alisha-Nehring	META_Signed-after-eFile	174	7.53%
5	AUTH_Amanda-Burg	META_Signed-after-eFile	63	5.33%
6	AUTH_Megan-Larison-DHS	META_Signed-after-eFile	42	4.23%
7	AUTH_Judith-Cole	META_Signed-after-eFile	80	3.14%
8	AUTH_Amanda-Burg	META_ESolutions-Signature	145	3.07%
...	...	...	...	...
28	AUTH_Megan-Larison-DHS	META_Created-after-eFile	17	0.12%

Top 10 Cross-Group Shared Row Hashes (≥2 groups, ≥365-day span):

Text Sample	Groups Present	Total Docs	Date Span (days)
'State of Minnesota'	8 of 8	641	2,888
'Filed in District Court'	8 of 8	609	3,104
'STATE OF MINNESOTA DISTRICT COURT'	8 of 8	347	3,020
'Judge of District Court'	8 of 8	333	3,020
'2'	8 of 8	327	3,104
'ORDER'	8 of 8	281	3,104
'Defendant.'	8 of 8	190	2,837
'COUNTY OF HENNEPIN FOURTH JUDICIAL DISTRICT'	7 of 8	279	—
'3'	7 of 8	256	3,101
'Plaintiff,'	7 of 8	179	2,855
'Defendant is presently incompetent to stand trial.'	4 of 8	51	2,990

Significance

The highest-Jaccard pair — AUTH_Alisha-Nehring and META_ESolutions-Signature at 18.63% — links a named-author group to a metadata-signature group. These groups share 960 distinct text row hashes, indicating that documents attributed to Alisha Nehring are produced within the same template infrastructure that generates ESolutions-signed documents. The second-highest pair (META_ESolutions-Signature / META_Signed-after-eFile at 17.66%) links two metadata-property groups whose document sets are textually nearly interchangeable.

The AUTH_Amanda-Burg / AUTH_Megan-Larison-DHS coupling (16.74%) is notable: two AUTH groups representing different named authors share more text per unit of vocabulary than most META-META pairs, suggesting their templates derive from the same or closely related source material.

Connecting Context

Part 1 found that AUTH_Alisha-Nehring has 29 documents and META_ESolutions-Signature has 134. The 960 shared row hashes between them exceed the total distinct row hash count within AUTH_Alisha-Nehring itself (1,628), indicating these groups share a substantial vocabulary — not merely a few header lines. Cross-referencing Part 1's multi-membership data: AUTH_Alisha-Nehring documents are not in META_ESolutions-Signature, making this a purely content-based coupling across structurally separate groups.

Finding 4: Stream Dictionary Fingerprints & Rendering Pipeline Signatures

Background

Page content streams in PDF files contain the binary rendering instructions that produce each page's visual output. Their SHA256 hashes (decoded_sha256 in hashpack_page_content_streams) serve as pipeline fingerprints: documents generated by the same software with the same settings produce identical stream hashes. High concentration of a single stream hash within a group indicates uniform rendering infrastructure. Cross-group stream hash sharing indicates shared rendering software or templates.

Findings

Stream Count Profile per Group:

Group	Docs w/ Streams	Total Stream Occurrences	Distinct Stream Hashes	Avg Streams / Doc	Median	Top-1 Hash Doc Count	Cross-Group Sharing
AUTH_Alisha-Nehring	29	249	164	20.5	20	4 / 29	shared (2 grps)
AUTH_Amanda-Burg	75	263	210	12.1	10	4 / 75	unique (1 grp)
AUTH_Judith-Cole	123	1,325	463	24.0	25	20 / 123	unique (1 grp)
AUTH_Megan-Larison-DHS	20	41	41	9.7	10	1 / 20	unique (1 grp)
META_Created-after-eFile	656	2,485	1,396	12.8	9	298 / 656	unique (1 grp)
META_ESolutions-Signature	134	1,111	864	25.1	25	5 / 134	unique (1 grp)
META_MSIP	34	203	171	16.7	16	26 / 34	unique (1 grp)
META_Signed-after-eFile	20	142	141	27.4	24	2 / 20	shared (3 grps)

Key observations from the stream analysis:

- META_Created-after-eFile: The dominant stream hash appears in 298 of 656 documents (45.4%) — nearly half the group. This single stream hash represents a common page rendering template. It is not shared with any other group, indicating a rendering pipeline specific to this cohort.
- META_MSIP: The dominant stream hash appears in 26 of 34 documents (76.5%) — the highest within-group concentration. This hash is also exclusively within META_MSIP and not shared with other groups, consistent with a distinct appellate-document rendering pipeline.
- AUTH_Judith-Cole: The top 5 stream hashes each appear in exactly 20 documents and are not shared with any other group. The pattern (5 dominant hashes × 20 documents each = 100 stream occurrences from 1,325 total) suggests structured template variants with consistent rendering per variant.
- AUTH_Megan-Larison-DHS: 41 stream occurrences across 20 documents — each stream hash appears in exactly 1 document (zero cross-document stream reuse). This is the only group with no stream-level reuse, indicating individually-rendered documents despite high row-level template reuse.
- META_Signed-after-eFile: Top stream hashes are shared across 3 groups — the only group showing significant cross-group stream sharing. This suggests these documents pass through a shared rendering layer.

Significance

Stream-level analysis reveals rendering pipeline separation: each group's dominant stream hashes are predominantly exclusive to that group, with the exception of META_Signed-after-eFile. The contrast between AUTH_Megan-Larison-DHS (zero stream reuse, high row reuse) and AUTH_Judith-Cole (systematic stream reuse, high row reuse) suggests

different document assembly approaches: Megan-Larison-DHS documents may be individually rendered from a shared text template, while Judith-Cole documents may use a small set of master rendering configurations.

Connecting Context

The META_MSIP stream concentration (76.5%) is consistent with Part 1's finding that META_MSIP documents are 79.4% MN Court of Appeals case type — a single appellate jurisdiction with its own document management infrastructure. The distinct stream fingerprint confirms a separate rendering pathway from the district-court groups.

Finding 5: Outline / Bookmark Hash Analysis

Background

PDF outlines (bookmarks) are structured navigation elements embedded in documents. When present, their titles and structure are stored in hashpack_outline_lines. Two documents sharing the same outline title text are structurally identical at the navigation layer, indicating they derive from the same document template with the same section structure. Groups with high outline coverage and low title diversity are template-driven.

Findings

Group	Total Docs	Docs with Outline	Coverage %	Total Outline Rows	Distinct Titles
META_MSIP	34	28	82.4%	35	5
AUTH_Alisha-Nehring	29	8	27.6%	16	2
AUTH_Judith-Cole	123	27	22.0%	107	5
META_Signed-after-eFile	20	4	20.0%	7	5
META_ESolutions-Signature	134	17	12.7%	38	16
META_Created-after-eFile	656	13	2.0%	25	11
AUTH_Megan-Larison-DHS	20	0	0.0%	0	0
AUTH_Amanda-Burg	75	0	0.0%	0	0

Top Outline Titles and Cross-Group Presence:

Title	Group	Docs	Groups Sharing This Title
'ORDER'	AUTH_Judith-Cole	27	6 of 8
'FINDINGS OF FACT'	AUTH_Judith-Cole	27	6 of 8
'State of Minnesota,)'	AUTH_Judith-Cole	27	4 of 8
'O R D E R'	META_MSIP	18	1 of 8 (exclusive)
'ORDER'	META_MSIP	6	6 of 8
'STATE OF MINNESOTA IN COURT OF APPEALS'	META_MSIP	9	1 of 8 (exclusive)
'ORDER'	AUTH_Alisha-Nehring	8	6 of 8
'FINDINGS OF FACT'	AUTH_Alisha-Nehring	8	6 of 8
'ORDER'	META_ESolutions-Signature	8	6 of 8

Significance

META_MSIP has the highest outline coverage at 82.4%, with only 5 distinct titles in 35 outline rows. Its two dominant exclusive titles — 'O R D E R' (formatted with spaces, unique to MSIP, 18 docs) and 'STATE OF MINNESOTA IN COURT OF APPEALS' (9 docs) — are hallmarks of Minnesota Court of Appeals document templates. These titles are not shared with any other group, confirming a structurally distinct appellate production pipeline.

The title pair 'ORDER' and 'FINDINGS OF FACT' appears across 6 of 8 groups — a structural template vocabulary shared by most district-court document types. AUTH_Judith-Cole uses this pair in every one of its 27 outline-bearing documents (100% consistency). AUTH_Amanda-Burg and AUTH_Megan-Larison-DHS have zero outline coverage — meaning these document types are produced without navigation structure, consistent with simple single-section form documents.

Connecting Context

The 'O R D E R' title format (with spaces) is exclusive to META_MSIP's 28 appellate documents — a formatting convention specific to Minnesota Court of Appeals document templates. Combined with Finding 4's distinct stream fingerprint and Part 1's finding of a common M365 tenant across three judicial branches, the outline data confirms META_MSIP represents a separate production infrastructure from all district-court groups.

Finding 6: Content Uniqueness Cards Per Group

Background

The Content Uniqueness Card synthesizes all hash-based metrics into a single per-group profile. It combines the clone rate (whole-document fingerprint), row-level template reuse (within-group row hashes), per-document uniqueness (fraction of each document's rows that are unique to it within its group), stream concentration, and outline coverage. The card identifies which group is most similar at the row-hash level (maximum Jaccard partner).

Findings

Group	Exact-Clone Rate %	Template Reuse %	Mean Doc Uniqueness %	Top-1 Row Hash Text	Max-Jaccard Partner	Outline %
AUTH_Alisha-Nehring	31.0%	67.6%	17.5%	'Plaintiff,'	META_ESolutions-Sig (18.63%)	27.6%
AUTH_Amanda-Burg	10.7%	91.1%	3.7%	'2'	AUTH_Megan-Larison (16.74%)	0.0%
AUTH_Judith-Cole	73.2%	91.2%	5.6%	'3'	META_Signed-after-eFile (3.14%)	22.0%
AUTH_Megan-Larison-DHS	0.0%	82.1%	7.7%	'By presentation of...'	AUTH_Amanda-Burg (16.74%)	0.0%
META_Created-after-eFile	26.7%	63.1%	21.7%	'State of Minnesota'	META_ESolutions-Sig (1.86%)	2.0%
META_ESolutions-Signature	27.6%	58.5%	29.1%	'State of Minnesota'	AUTH_Alisha-Nehring (18.63%)	12.7%
META_MSIP	0.0%	30.2%	58.8%	'2'	META_ESolutions-Sig (2.87%)	82.4%
META_Signed-after-eFile	0.0%	25.5%	63.8%	'Filed in District Court'	META_ESolutions-Sig (17.66%)	20.0%

Significance

Three distinct content profiles emerge:

- High-Homogeneity (Template Factories): AUTH_Judith-Cole and AUTH_Amanda-Burg. These groups have >90% row-level template reuse, mean per-document uniqueness below 6%, and multiple exact-content clones. They represent document production systems operating from highly constrained master templates.
- Mid-Homogeneity (Structured but Variant): AUTH_Megan-Larison-DHS, AUTH_Alisha-Nehring, META_Created-after-eFile, META_ESolutions-Signature. Template reuse ranges 58–82%, uniqueness 17–29%. These groups have template foundations but more per-document variability in the case-specific content layers.
- Low-Homogeneity (Content-Diverse): META_MSIP and META_Signed-after-eFile. Exact-clone rate 0%, template reuse 25–30%, mean uniqueness 59–64%. Each document in these groups is textually distinct from its peers — the metadata properties that qualify them (MSIP label, post-eFile signature timestamp) are not associated with templated content.

LIMITATIONS & ALTERNATIVE EXPLANATIONS

- Row-level text hashing after PDF text extraction may produce identical hashes for semantically distinct content that happens to extract identically (e.g., single-character rows like '2' or '3'). Page-number rows are included in the row hash pool and inflate template reuse counts for groups with high page-number consistency.
- The outline analysis uses exact title text (not a separate hash column), meaning minor formatting differences in capitalization or spacing would produce different records even for structurally equivalent titles. The 'O R D E R' vs 'ORDER' distinction observed in META_MSIP is real and meaningful, but intermediate cases may exist.
- Cross-group Jaccard values are computed from the full row-hash vocabulary including legal boilerplate. The 7 universal row hashes (present in all 8 groups) inflate pairwise similarity across all group combinations. Filtering these universal rows before computing Jaccard would produce lower but more discriminating values.
- Stream hash analysis uses hashpack_page_content_streams which reflects the PDF's decoded page content streams. Documents with identical visual output but different compression settings, font encoding, or annotation layers may produce different stream hashes. The converse — identical stream hashes from different documents — is strong evidence of common origin.
- Alternative explanation for high template reuse in AUTH groups: Document management systems commonly store form templates and populate only case-specific fields, resulting in high structural homogeneity. The measured reuse rates are consistent with this workflow. The extent to which the specific reuse rates are anomalous relative to other court document corpora cannot be determined from this dataset alone.
- The corpus is limited to documents tagged in the 8 doc_set groups (1,043 of 4,251 total PDFs). The broader corpus may show different baseline template reuse rates. Cross-group Jaccard values would shift with different cohort boundaries.

RECOMMENDATIONS FOR INDEPENDENT VERIFICATION

- Reproduce the Anchor CTE using the confirmed column name 'value' in children_doc_set. Verify all 8 group counts match this report before running downstream analyses.
- For fp_text_sha256 clone verification: extract the top AUTH_Judith-Cole clone group (20 documents sharing one fp_text_sha256) and compare the actual PDF text side-by-side. Obtain their original filings from the court's internal document management system and compare against MCRO-served versions.
- For cross-group Jaccard reproduction: execute Appendix A5 using the same Anchor CTE. The cross-group Jaccard computation is $O(N^2)$ on distinct row hash sets — for the full corpus this is tractable. Consider filtering universal rows (those appearing in all 8 groups) before computing group-specific similarity.
- For stream pipeline verification: submit the top META_Created-after-eFile stream hash (present in 298 documents) and the top META_MSIP stream hash (present in 26 documents) to a PDF forensics tool capable of decoding page content streams. Determine whether these are standard rendering artifacts or custom-generated content.
- For outline coverage verification: re-run hashpack_outline_lines queries using the confirmed column names. The absence of a hash column means outline comparison relies on exact text matching — verify that the 'O R D E R' (spaced) format is consistently applied across all 18 META_MSIP documents showing that title.
- For AUTH_Judith-Cole clone investigation: the 17 clone groups within 123 documents suggest 17 distinct template variants. Obtain the filing dates and case IDs for each clone group to determine whether clones are filed in temporal clusters or distributed chronologically — temporal clustering would indicate batch document generation events.

APPENDIX | SQL QUERIES

All queries are SELECT-only. Column name discoveries (value, text_clean, n_pages, fp_streams_other_stream_count) are documented inline.

A1 — Anchor CTE (Identical to Part 1)

```
WITH doc_set_cohort AS (   SELECT     cds.evidence_uid,     cds.value AS doc_set,   -- NOTE:
column is 'value', not 'doc_set'     c.case_uid,     c.cluster_id,     c.cluster_name,
c.filing_type,     c.filing_date,     c.case_type,     c.filename,     c.me_filename,
c.doc_sha256   FROM children_doc_set cds   JOIN children c ON c.evidence_uid = cds.evidence_uid
WHERE cds.value IN (     'META_ESolutions-Signature',     'META_Created-after-eFile',
'META_MSIP',     'META_Signed-after-eFile',     'AUTH_Alisha-Nehring',     'AUTH_Amanda-Burg',
'AUTH_Judith-Cole',     'AUTH_Megan-Larison-DHS'   ) )
```

A2 — Cohort Verification (COUNT per group)

```
-- Run before any analysis to verify Part 1 cohort counts SELECT    cds.value AS doc_set,
COUNT(cds.evidence_uid) AS per_group_rows,   COUNT(DISTINCT cds.evidence_uid) AS distinct_docs
FROM children_doc_set cds WHERE cds.value IN (   'META_ESolutions-Signature','META_Created-after-
eFile','META_MSIP',   'META_Signed-after-eFile','AUTH_Alisha-Nehring','AUTH_Amanda-Burg',
'AUTH_Judith-Cole','AUTH_Megan-Larison-DHS' ) GROUP BY cds.value ORDER BY cds.value;
```

A3 — Analysis 1: Hashpack Coverage & Exact-Clone Rates

```
WITH doc_set_cohort AS ( [... Anchor CTE from A1 ...] ), -- fp_text_sha256 per group fp_counts AS
(   SELECT dsc.doc_set, h.fp_text_sha256, COUNT(*) AS cnt   FROM doc_set_cohort dsc   JOIN
hashpack h ON h.evidence_uid = dsc.evidence_uid   WHERE h.fp_text_sha256 IS NOT NULL   GROUP BY
dsc.doc_set, h.fp_text_sha256 ), clone_docs AS (   SELECT dsc.doc_set, dsc.evidence_uid,     CASE
WHEN fp_counts.cnt > 1 THEN 1 ELSE 0 END AS is_clone   FROM doc_set_cohort dsc   JOIN hashpack h
ON h.evidence_uid = dsc.evidence_uid   JOIN fp_counts ON fp_counts.doc_set = dsc.doc_set      AND
fp_counts.fp_text_sha256 = h.fp_text_sha256 ) SELECT doc_set,   COUNT(*) AS total_docs,
SUM(is_clone) AS cloned_docs,   ROUND(100.0 * SUM(is_clone) / COUNT(*), 1) AS exact_clone_rate_pct
FROM clone_docs GROUP BY doc_set ORDER BY exact_clone_rate_pct DESC;  -- Clone group size
distribution WITH ... fp_groups AS (   SELECT dsc.doc_set, h.fp_text_sha256, COUNT(*) AS
clone_group_size   FROM doc_set_cohort dsc   JOIN hashpack h ON h.evidence_uid = dsc.evidence_uid
WHERE h.fp_text_sha256 IS NOT NULL   GROUP BY dsc.doc_set, h.fp_text_sha256   HAVING COUNT(*) >
1 ) SELECT doc_set, COUNT(*) AS num_clone_groups, MAX(clone_group_size) AS max_clone_size FROM
fp_groups GROUP BY doc_set;
```

A4 — Analysis 2: Row-Level Collision Rates (Within-Group)

```
WITH doc_set_cohort AS ( [... Anchor CTE ...] ), -- Within-group row uniqueness row_stats AS
(   SELECT dsc.doc_set,     COUNT(*) AS total_rows,     COUNT(DISTINCT hr.row_sha256) AS
distinct_row_sha256   FROM doc_set_cohort dsc   JOIN hashpack_rows hr ON hr.evidence_uid =
dsc.evidence_uid   -- NOTE: column is 'text_clean' not 'row_text'   GROUP BY dsc.doc_set ) SELECT
doc_set, total_rows, distinct_row_sha256,   ROUND(100.0 * distinct_row_sha256 / total_rows, 1) AS
uniqueness_pct,   ROUND(100.0 * (total_rows - distinct_row_sha256) / total_rows, 1) AS
template_reuse_pct FROM row_stats ORDER BY template_reuse_pct DESC;  -- Tier distribution WITH ...
row_doc_counts AS (   SELECT dsc.doc_set, hr.row_sha256, COUNT(DISTINCT hr.evidence_uid) AS
doc_count   FROM doc_set_cohort dsc   JOIN hashpack_rows hr ON hr.evidence_uid = dsc.evidence_uid
GROUP BY dsc.doc_set, hr.row_sha256 ), tiered AS (   SELECT doc_set, row_sha256,     CASE WHEN
doc_count = 1 THEN 'unique (1 doc)'          WHEN doc_count BETWEEN 2 AND 5 THEN '2-5 docs'
WHEN doc_count BETWEEN 6 AND 20 THEN '6-20 docs'          ELSE '20+ docs (boilerplate)' END AS
tier   FROM row_doc_counts ) SELECT doc_set, tier, COUNT(*) AS distinct_row_hashes FROM tiered
GROUP BY doc_set, tier ORDER BY doc_set, tier;  -- Per-document uniqueness score WITH ...
group_row_freq AS (   SELECT dsc.doc_set, hr.row_sha256, COUNT(DISTINCT hr.evidence_uid) AS
doc_freq   FROM doc_set_cohort dsc   JOIN hashpack_rows hr ON hr.evidence_uid = dsc.evidence_uid
GROUP BY dsc.doc_set, hr.row_sha256 ), doc_unique AS (   SELECT dsc.doc_set, hr.evidence_uid,
SUM(CASE WHEN grf.doc_freq = 1 THEN 1 ELSE 0 END) AS unique_rows,     COUNT(*) AS total_rows
FROM doc_set_cohort dsc   JOIN hashpack_rows hr ON hr.evidence_uid = dsc.evidence_uid   JOIN
group_row_freq grf ON grf.doc_set = dsc.doc_set AND grf.row_sha256 = hr.row_sha256   GROUP BY
dsc.doc_set, hr.evidence_uid ) SELECT doc_set,   ROUND(AVG(100.0 * unique_rows /
total_rows)::numeric, 1) AS mean_uniqueness_pct,   ROUND(PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER
BY 100.0*unique_rows/total_rows)::numeric,1) AS median_pct FROM doc_unique GROUP BY doc_set ORDER
BY mean_uniqueness_pct;
```

A5 — Analysis 3: Cross-Group Jaccard Similarity Matrix

```
WITH doc_set_cohort AS ( [... Anchor CTE ...] ), group_row_hashes AS (   SELECT DISTINCT
dsc.doc_set, hr.row_sha256   FROM doc_set_cohort dsc   JOIN hashpack_rows hr ON hr.evidence_uid =
dsc.evidence_uid ), group_sizes AS (   SELECT doc_set, COUNT(DISTINCT row_sha256) AS n_hashes
FROM group_row_hashes GROUP BY doc_set ), pairs AS (   SELECT a.doc_set AS group_a, b.doc_set AS
group_b, COUNT(*) AS intersection_size   FROM group_row_hashes a   JOIN group_row_hashes b ON
a.row_sha256 = b.row_sha256 AND a.doc_set < b.doc_set   GROUP BY a.doc_set, b.doc_set ) SELECT
p.group_a, p.group_b, p.intersection_size,   sa.n_hashes AS size_a, sb.n_hashes AS size_b,
ROUND(100.0 * p.intersection_size / (sa.n_hashes + sb.n_hashes - p.intersection_size), 2) AS
jaccard_pct FROM pairs p JOIN group_sizes sa ON sa.doc_set = p.group_a JOIN group_sizes sb ON
sb.doc_set = p.group_b ORDER BY jaccard_pct DESC;  -- Cross-group shared row hashes with long date
span WITH ... cross_group AS (   SELECT hr.row_sha256,     COUNT(DISTINCT dsc.doc_set) AS
group_count,     COUNT(DISTINCT hr.evidence_uid) AS doc_count,     MIN(dsc.filing_date) AS
first_date, MAX(dsc.filing_date) AS last_date,     MAX(dsc.filing_date) - MIN(dsc.filing_date) AS
span_days,     MAX(LEFT(hr.text_clean, 80)) AS text_sample   FROM doc_set_cohort dsc   JOIN
hashpack_rows hr ON hr.evidence_uid = dsc.evidence_uid   GROUP BY hr.row_sha256   HAVING
COUNT(DISTINCT dsc.doc_set) > 1     AND MAX(dsc.filing_date) - MIN(dsc.filing_date) >= 365 )
SELECT * FROM cross_group ORDER BY span_days DESC LIMIT 20;
```

A6 — Analysis 4: Stream Dictionary Fingerprints

```
WITH doc_set_cohort AS ( [... Anchor CTE ...] ), -- Stream stats per group stream_stats AS
(   SELECT dsc.doc_set,     COUNT(DISTINCT dsc.evidence_uid) AS docs_with_streams,
COUNT(hpcs.decoded_sha256) AS total_occurrences,     COUNT(DISTINCT hpcs.decoded_sha256) AS
distinct_stream_hashes   FROM doc_set_cohort dsc   JOIN hashpack_page_content_streams hpcs ON
hpcs.evidence_uid = dsc.evidence_uid   GROUP BY dsc.doc_set ) SELECT * FROM stream_stats ORDER BY
doc_set;  -- Top 5 stream hashes per group with cross-group sharing WITH ... stream_freq AS
(   SELECT dsc.doc_set, hpcs.decoded_sha256,     COUNT(*) AS occurrences,     COUNT(DISTINCT
hpcs.evidence_uid) AS doc_count   FROM doc_set_cohort dsc   JOIN hashpack_page_content_streams
hpcs ON hpcs.evidence_uid = dsc.evidence_uid   GROUP BY dsc.doc_set, hpcs.decoded_sha256 ), ranked
AS (   SELECT *, ROW_NUMBER() OVER (PARTITION BY doc_set ORDER BY doc_count DESC) AS rn   FROM
stream_freq ), cross_group AS (   SELECT decoded_sha256, COUNT(DISTINCT doc_set) AS n_groups FROM
stream_freq GROUP BY decoded_sha256 ) SELECT r.doc_set, r.rn, r.decoded_sha256, r.doc_count,
cg.n_groups FROM ranked r JOIN cross_group cg ON cg.decoded_sha256 = r.decoded_sha256 WHERE r.rn
<= 5 ORDER BY r.doc_set, r.rn;  -- Stream count per doc per group (using
fp_streams_other_stream_count from hashpack) SELECT dsc.doc_set,
ROUND(AVG(h.fp_streams_other_stream_count)::numeric, 1) AS avg_streams,
MIN(h.fp_streams_other_stream_count), MAX(h.fp_streams_other_stream_count),   PERCENTILE_CONT(0.5)
WITHIN GROUP (ORDER BY h.fp_streams_other_stream_count) AS median_streams FROM doc_set_cohort dsc
JOIN hashpack h ON h.evidence_uid = dsc.evidence_uid GROUP BY dsc.doc_set ORDER BY avg_streams
DESC;
```

A7 — Analysis 5: Outline / Bookmark Analysis

```
WITH doc_set_cohort AS ( [... Anchor CTE ...] ) -- NOTE: hashpack_outline_lines has no hash
column; title comparison uses text -- Columns: evidence_uid, outline_index,
outline_line_json__outline_title,  --          outline_line_json__outline_level,
outline_line_json__outline_page_1based  -- Coverage per group SELECT dsc.doc_set,   COUNT(DISTINCT
dsc.evidence_uid) AS total_docs,   COUNT(DISTINCT hol.evidence_uid) AS docs_with_outline,
ROUND(100.0 * COUNT(DISTINCT hol.evidence_uid) / COUNT(DISTINCT dsc.evidence_uid), 1) AS
coverage_pct,   COUNT(hol.hashpack_outline_lines_row_id) AS total_rows,   COUNT(DISTINCT
hol.outline_line_json__outline_title) AS distinct_titles FROM doc_set_cohort dsc LEFT JOIN
hashpack_outline_lines hol ON hol.evidence_uid = dsc.evidence_uid GROUP BY dsc.doc_set ORDER BY
coverage_pct DESC;  -- Top titles per group with cross-group sharing WITH title_freq AS (   SELECT
dsc.doc_set, hol.outline_line_json__outline_title AS title,     COUNT(DISTINCT hol.evidence_uid)
AS doc_count   FROM doc_set_cohort dsc   JOIN hashpack_outline_lines hol ON hol.evidence_uid =
dsc.evidence_uid   GROUP BY dsc.doc_set, hol.outline_line_json__outline_title ), cross_group AS (
SELECT hol.outline_line_json__outline_title AS title, COUNT(DISTINCT dsc.doc_set) AS n_groups
FROM doc_set_cohort dsc   JOIN hashpack_outline_lines hol ON hol.evidence_uid = dsc.evidence_uid
GROUP BY title ), ranked AS (   SELECT *, ROW_NUMBER() OVER (PARTITION BY doc_set ORDER BY
doc_count DESC) AS rn FROM title_freq ) SELECT r.doc_set, r.title, r.doc_count, cg.n_groups FROM
ranked r JOIN cross_group cg ON cg.title = r.title WHERE r.rn <= 3 ORDER BY r.doc_set, r.rn;
```

A8 — Python: Jaccard Heatmap (matplotlib)

```
import matplotlib.pyplot as plt import matplotlib.colors as mcolors import numpy as np  groups =
[     "AUTH_Alisha-Nehring", "AUTH_Amanda-Burg", "AUTH_Judith-Cole",     "AUTH_Megan-Larison-DHS",
"META_Created-after-eFile",     "META_ESolutions-Signature", "META_MSIP", "META_Signed-after-eFile" ]
short = ["AN", "AB", "JC", "ML", "CE", "ES", "MS", "SF"]  # Jaccard values from DB (upper triangle,
symmetric) pairs = {     (0,1):1.41,(0,2):1.07,(0,3):0.95,(0,4):1.02,(0,5):18.63,(0,6):1.09,(0,7):7.53,
(1,2):0.51,(1,3):16.74,(1,4):0.17,(1,5):3.07,(1,6):0.44,(1,7):5.33,     (2,3):0.52,(2,4):1.73,(2,5):2.25,
(2,6):0.41,(2,7):3.14,     (3,4):0.12,(3,5):1.06,(3,6):0.40,(3,7):4.23,     (4,5):1.86,(4,6):0.45,
(4,7):1.86,     (5,6):2.87,(5,7):17.66,     (6,7):0.91 } mat = np.zeros((8,8)) for (i,j),v in
pairs.items():     mat[i,j] = v; mat[j,i] = v  fig, ax = plt.subplots(figsize=(10,8)) cmap =
plt.cm.YlOrRd im = ax.imshow(mat, cmap=cmap, vmin=0, vmax=20) ax.set_xticks(range(8));
ax.set_yticks(range(8)) ax.set_xticklabels(short, fontsize=10) ax.set_yticklabels(short, fontsize=10) for
i in range(8):     for j in range(8):         if i != j:             ax.text(j, i, f"{mat[i,j]:.1f}",
ha='center', va='center', fontsize=8,                     color='white' if mat[i,j]>12 else 'black')
else:             ax.text(j, i, "—", ha='center', va='center', fontsize=9, color='gray') plt.colorbar(im,
ax=ax, label='Jaccard Similarity %') ax.set_title('Cross-Group Row Hash Jaccard Similarity Matrix\n(MCRO
Doc-Set Forensics Part 2)',              fontsize=12, fontweight='bold') plt.tight_layout()
plt.savefig('jaccard_heatmap.png', dpi=150) plt.show()
```

POSSIBLE EXPLANATIONS (Unresolved; Not Proven by This Report)

The following speculative explanations are offered without endorsement. None is confirmed by the available data. Independent investigation would be required to establish or rule out any of these hypotheses.

- The high exact-clone rate within AUTH_Judith-Cole (73.2%) may reflect a court document management system that stamps a stored form template into an evidence record without inserting case-specific content beyond the case caption — producing textually identical orders for different defendants.
- The AUTH_Alisha-Nehring / META_ESolutions-Signature coupling (Jaccard 18.63%) may indicate that the Nehring-attributed documents pass through the same Tyler Technologies / ESolutions pipeline that generates the ESolutions-signed documents — a shared workflow rather than a shared authorship.
- The META_MSIP group's distinct stream pipeline and exclusive appellate outline formatting ('O R D E R' with spaces) may reflect a separate document generation system operated within the same M365 organizational tenant identified in the Initial Forensic Analysis Report Finding 4 — consistent with appellate-branch document production infrastructure.
- AUTH_Amanda-Burg / AUTH_Megan-Larison-DHS Jaccard coupling (16.74%) may result from these two examiner roles using the same DHS competency evaluation form template, which would be expected given they are both DHS forensic evaluators operating under the same institutional framework.
- The zero-reuse stream layer in AUTH_Megan-Larison-DHS despite high row-level reuse may indicate these documents are printed-to-PDF rather than template-generated — a workflow where the same form text is entered into a word processor and saved as PDF, producing unique stream binaries despite identical form content.