

FORENSIC ANALYSIS REPORT
Minnesota Court Records Online (MCRO)

MCRO | Doc-Set Group Forensics

Part 5 of 5: Language & Terminology Analysis

8 Doc-Set Groups · 1,043 Unique Documents · Series Capstone Report

Field	Value
Series	Part 5 of 5 — Doc-Set Group Forensics (Final / Capstone)
Report Date	March 2026
Scope	Language & terminology analysis via children_doc_strings across 8 doc_set groups
Cohort	1,091 per-group assignments 1,043 unique documents 4 META + 4 AUTH groups
Database	PostgreSQL (Supabase ibfmjtwahkwqzcmeyqii) — READ-ONLY Queried March 2026
Prepared by	Matthew David Guertin, Pro Se Defendant Case 27-CR-23-1886
Court	Fourth Judicial District Court, Hennepin County, Minnesota

IMPORTANT DISCLAIMER

This report is Part 5 of 5 — the capstone of the Doc-Set Group Forensics series. Parts 1–4 established the cohort census (Part 1), content hash fingerprints (Part 2), PDF internal object inventories (Part 3), and digital signatures & persons profiles (Part 4). Part 5 completes the multi-dimensional analysis by mapping the linguistic content of each group via the `children_doc_strings` table.

All statistics were freshly computed via SELECT-only queries against the live PostgreSQL database in March 2026. No statistics are reused from prior reports or external files. This report does not assert conclusions about guilt, fabrication, or criminal intent. Named persons appearing as `string_norm` values are reported as database facts (they appear in the extracted text of court documents), not as accusations.

Findings represent measurable patterns in document terminology and vocabulary composition. Each finding requires independent expert examination before use in formal legal proceedings. It is strongly recommended that a credentialed digital forensics examiner (CFCE, EnCE, or equivalent) independently verify these findings.

NOTE ON COLUMN NAMING: The `children_doc_set` table uses the column name `'value'` (not `'doc_set'`). The `children_doc_strings` table uses `'string_source_url'` (not `'source_url'`). All SQL uses confirmed live column names.

EXECUTIVE SUMMARY

- Cohort confirmed: 1,091 per-group assignments / 1,043 unique documents — exact match with Part 1.
- Language coverage ranges from 56.1% (META_Created-after-eFile) to 100% (all AUTH groups + META_MSIP). The 288 uncovered META_Created-after-eFile documents represent the largest gap in the linguistic corpus.
- AUTH_Amanda-Burg: Only 29 distinct terms — the smallest vocabulary. Its top-10 terms cover 78.0% of all rows, and 26.8% of rows are a single term (MN Department of Human Services).
- META_Created-after-eFile: 291 distinct terms and 98 group-exclusive terms (33.7% exclusivity) — by far the broadest and most distinctive vocabulary. Exclusive terms include 'Google', 'infiniset', 'USPTO', 'CIA', and 'synthetic' — vocabulary absent from every other group.
- 9 universal terms appear in all 8 groups: 'Rule 20 AND Rule 20.01', 'services', 'health', 'care', 'facility', 'mental health', 'participate', 'competency to proceed', and 'By the court'.
- Highest vocabulary similarity: AUTH_Amanda-Burg ↔ AUTH_Megan-Larison-DHS (Jaccard = 74.2%). Both are DHS-originated document groups with nearly identical template vocabulary.
- Lowest vocabulary similarity: AUTH_Megan-Larison-DHS ↔ META_Created-after-eFile (Jaccard = 7.8%). The linguistically most distant pair in the corpus.
- META-only terms (179) vastly outnumber AUTH-only terms (11). META groups collectively own a vocabulary 16× larger than AUTH-exclusive terminology.
- 100% of cohort string rows carry a non-null `string_source_url`, providing direct page-level hyperlinks to source documents for every extracted term.
- Zero `string_id` collisions: every `string_id` maps to exactly one `string_norm`. Two terms ('disorder', 'impairment') appear under dual `string_group_type` categories by design.

TABLE ES-1: Language Coverage & Key Metrics Per Group

Group	Type	Docs	w/Str	Cov%	Rows	Terms	Excl	Top SGT
META_Created-after-eFile	MET A	656	368	56.1%	9,218	291	98	experts (26.0%)
META_ESolutions-Signature	MET A	134	132	98.5%	4,856	261	19	experts (18.5%)
META_MSIP	MET A	34	34	100%	446	106	8	protect (25.6%)
META_Signed-after-eFile	MET A	20	19	95.0%	345	79	0	experts (23.2%)
AUTH_Alisha-Nehring	AUTH	29	29	100%	3,738	203	4	dsm5 (19.7%)
AUTH_Amanda-Burg	AUTH	75	75	100%	3,637	29	1	itinerary (35.8%)
AUTH_Judith-Cole	AUTH	123	123	100%	10,655	80	4	experts (24.2%)
AUTH_Megan-Larison-DHS	AUTH	20	20	100%	750	25	0	itinerary (34.0%)

► KEY METRICS — NUMBERS THAT DEMAND ATTENTION

- META_Created-after-eFile: 98 exclusive terms (33.7%) including 'Google', 'infiniset', 'USPTO', 'CIA', 'LinkedIn', 'synthetic', 'metadata', 'compel discovery' — vocabulary found in zero other groups.
- AUTH_Amanda-Burg + AUTH_Megan-Larison-DHS: Jaccard = 74.2% — nearly three-quarters vocabulary overlap from only 29 and 25 distinct terms respectively. Both DHS-originated.
- AUTH_Alisha-Nehring ↔ META_ESolutions-Signature: Jaccard = 70.6% — the strongest META ↔ AUTH cross-boundary vocabulary link, despite different classification criteria.
- AUTH_Judith-Cole: 10,655 string rows across 123 documents (avg 86.6 rows/doc) — the most terminology-dense group. Lisa Janzen appears in 105 of 123 documents (85.4%).
- META_MSIP: 8 exclusive terms include 'Rebecca Cole', 'Zachary Cole', 'cole v cole' (person category), 'controlling', and 'constitutional violations'.
- 179 META-only terms vs. 11 AUTH-only terms: a 16:1 asymmetry in class-exclusive vocabulary.

METHODS

This report analyzes the linguistic content of the 8 doc_set groups by joining the Anchor CTE cohort (identical to Parts 1–4) to the children_doc_strings table on evidence_uid. The children_doc_strings table contains 181,556 rows corpus-wide, each representing an extracted and classified text string from a PDF document.

String Identity Trio

Each string row carries three identity fields: string_norm (the normalized canonical form — the primary vocabulary unit), string_raw (the original extracted text), and string_id (a composite identifier encoding the string_group_type and SHA256 hash). This trio enables normalization variance analysis and collision detection. The string_id format is str:{group_type}:sha256:{hash}, meaning the same normalized text classified under two different semantic categories receives two distinct string_ids by design.

Semantic Taxonomy (string_group_type)

Each string row is classified into a semantic category via the string_group_type column. Categories observed in the cohort include: experts (competency evaluation language), protect (court authority / protective language), dsm5 (clinical/diagnostic terminology), indefinite (commitment/detention language), itinerary (facility/transport references), control (behavioral control terms), competency (competency-specific terms), examiner (named forensic examiners), officer (judicial officers), attorney (named attorneys), reprogram (restoration terminology), mirror (credibility/opinion terms), truth (defendant self-referential claims), fraud (system integrity terms), laws (statutory references), monitoring (supervision terms), person (non-officer named persons), programs (program names), and liason (liaison references).

Multi-Group Membership

Per the Part 1 protocol, a term appearing in a document that belongs to multiple groups counts toward all its groups. This is consistent with the additive/non-exclusive counting approach used throughout the series. This is noted wherever term counts are presented.

Tables and Views Used

Primary: children_doc_set (cohort definition, column 'value'), children (document metadata), children_doc_strings (extracted terminology, column 'string_source_url'). Secondary: hashpack_rows (row-level SHA256 for cross-reference with Part 2). No signature, object, or XMP metadata tables were used.

Vocabulary Similarity

Cross-group vocabulary similarity is measured via Jaccard index: $J(A,B) = |A \cap B| / |A \cup B|$, computed over the distinct string_norm vocabularies of each group pair. The resulting 8×8 matrix quantifies linguistic overlap independent of term frequency.

FINDINGS

Finding 1: children_doc_strings Coverage Per Group

Background

Not all documents in the cohort necessarily have extracted terminology in children_doc_strings. Coverage depends on the text extraction pipeline and the document's content structure. Before analyzing vocabulary, the coverage rate per group must be established to contextualize all subsequent findings.

Findings

Group	Docs	w/Str	Cov%	Rows	Terms	SGTs	Avg R/D	Median R/D
META_Created-after-eFile	656	368	56.1%	9,218	291	18	25.0	2
META_ESolutions-Signature	134	132	98.5%	4,856	261	17	36.8	7
META_MSIP	34	34	100%	446	106	16	13.1	5
META_Signed-after-eFile	20	19	95.0%	345	79	15	18.2	8
AUTH_Alisha-Nehring	29	29	100%	3,738	203	14	128.9	80
AUTH_Amanda-Burg	75	75	100%	3,637	29	10	48.5	51
AUTH_Judith-Cole	123	123	100%	10,655	80	15	86.6	94
AUTH_Megan-Larison-DHS	20	20	100%	750	25	10	37.5	39

TABLE F1-1: Coverage overview. Cov% = % of group docs with ≥ 1 string row. Avg R/D and Median R/D = rows per document for docs with string data. SGTs = distinct string_group_type categories.

Significance

META_Created-after-eFile has the lowest coverage at 56.1% (288 documents without string data). Since this is the largest group (656 docs), this gap affects 27.6% of all unique cohort documents. All four AUTH groups achieve 100% coverage, reflecting their origin as authorship-attributed documents with substantive text content. AUTH_Alisha-Nehring has the highest text density (avg 128.9 rows/doc, median 80), while META_Created-after-eFile has a median of just 2 rows/doc — indicating that many of its documents with string data contain only minimal extracted terminology.

Connecting Context

Part 1 documented META_Created-after-eFile as containing the broadest range of filing types (including procedural documents like Notice of Case Reassignment that may contain minimal extractable clinical terminology). The 56.1% coverage is consistent with a group defined by a metadata property (XMP creation date post-dating e-file date) that captures documents regardless of textual content. See Appendix A2.

Finding 2: Term Inventory & Concentration Per Group

Background

Each group's vocabulary profile reveals whether its documents are terminologically diverse or dominated by a small set of repeated terms. Concentration metrics (top-1 and top-10 coverage) measure how much of a group's total string rows are attributable to its highest-frequency terms.

Findings

Group	Distinct Terms	Total Rows	Top-1 Term	Top-1 %	Top-10 %
META_Created-after-eFile	291	9,218	services	8.8%	45.3%
META_ESolutions-Sig	261	4,856	Rule 20 AND Rule 20.01	7.3%	34.1%
META_MSIP	106	446	duty	10.1%	47.5%
META_Signed-after-eFile	79	345	Rule 20 AND Rule 20.01	11.3%	52.8%
AUTH_Alisha-Nehring	203	3,738	Rule 20 AND Rule 20.01	5.3%	33.1%
AUTH_Amanda-Burg	29	3,637	MN Dept of Human Services	26.8%	78.0%
AUTH_Judith-Cole	80	10,655	commitment	10.8%	62.3%
AUTH_Megan-Larison-DHS	25	750	MN Dept of Human Services	25.2%	80.5%

Significance

AUTH_Amanda-Burg (29 terms) and AUTH_Megan-Larison-DHS (25 terms) have the smallest vocabularies with the highest concentration: their top-10 terms account for 78.0% and 80.5% of all rows respectively. Both are dominated by 'MN Department of Human Services' as the single most frequent term. This pattern is consistent with DHS-originated template documents with standardized language.

AUTH_Alisha-Nehring (203 terms) and META_ESolutions-Signature (261 terms) show the broadest term diversity with the lowest top-10 concentration (33.1% and 34.1%), indicating terminologically rich, non-template documents.

META_MSIP's top term is 'duty' (protect category, 45 occurrences in 17 docs) — a term reflecting appellate judicial language distinct from the clinical/competency vocabulary dominating other groups.

Connecting Context

Part 1 showed AUTH_Amanda-Burg was 54.7% 'Proposed Order or Document' filing type and AUTH_Judith-Cole was 85.4% 'Findings and Order'. The high concentration at the term level mirrors the filing-type homogeneity identified at the document level. See Appendix A3.

Finding 3: Exclusive vs Shared Terms

Background

A term is 'exclusive' to a group if it appears in ≥ 1 document in that group and in zero documents in any other group. Exclusive terms define what makes a group linguistically unique. Shared terms measure common ground across the corpus.

Findings

Group	Total Terms	Exclusive	Shared	Excl %
META_Created-after-eFile	291	98	193	33.7%
META_ESolutions-Signature	261	19	242	7.3%
META_MSIP	106	8	98	7.5%
META_Signed-after-eFile	79	0	79	0.0%
AUTH_Alisha-Nehring	203	4	199	2.0%
AUTH_Amanda-Burg	29	1	28	3.4%
AUTH_Judith-Cole	80	4	76	5.0%
AUTH_Megan-Larison-DHS	25	0	25	0.0%

META_Created-after-eFile Exclusive Terms (Top 15 by occurrence):

Term	SGT Category	Rows	Docs
Google	truth	81	72
infiniset	truth	69	7
compel discovery	fraud	47	3
synthetic	fraud	43	2
invention	truth	23	7
CIA	truth	22	4
C.I.A.	truth	22	4
LinkedIn	truth	22	5
military	truth	20	6
USPTO	truth	18	6
metadata	fraud	16	6
narrative	fraud	15	4
powerful people	fraud	14	6
financial	truth	13	4
bad bunny	truth	12	4

META_MSIP Exclusive Terms:

Term	SGT Category	Rows	Docs
Rebecca Cole	person	8	3
MN Stat. 61142 competency motion procedures	laws	5	3
controlling	control	3	1
Zachary Cole	person	3	3
cole v cole	person	1	1
illogical	control	1	1
Elizabeth Murphy	attorney	1	1
constitutional violations	fraud	1	1

Vocabulary Boundary Summary:

Category	Term Count	Description
Universal (all 8 groups)	9	Rule 20 AND Rule 20.01, services, health, care, facility, mental health, participate, competency to proceed, By the court
META-only terms	179	Terms appearing only in META groups, absent from all AUTH groups
AUTH-only terms	11	Terms appearing only in AUTH groups, absent from all META groups
Cross-boundary (META+AUTH)	207	Terms shared across the META/AUTH divide

Significance

META_Created-after-eFile's 98 exclusive terms are overwhelmingly classified as 'truth' (defendant self-referential claims like Google, LinkedIn, USPTO, CIA) and 'fraud' (system integrity terms like synthetic, compel discovery, metadata, narrative). No other group contains any of these terms. This vocabulary is structurally distinct from the clinical/legal terminology that dominates the other 7 groups.

META_MSIP's 8 exclusive terms include three 'person'-category entries (Rebecca Cole, Zachary Cole, cole v cole) that appear nowhere else in the cohort, plus 'constitutional violations' — language specific to the appellate filings that constitute this group. META_Signed-after-eFile and AUTH_Megan-Larison-DHS have zero exclusive terms — their entire vocabulary is a subset of other groups' language.

Connecting Context

Part 2 found that META_Created-after-eFile had the highest proportion of unique row hashes. The present finding extends that pattern to the term level: this group's documents contain not just unique text blocks but unique vocabulary absent from the broader corpus. See Appendix A4.

Finding 4: Cross-Group Vocabulary Similarity (Jaccard Matrix)

Background

The Jaccard index measures the proportion of shared vocabulary between two groups relative to their combined vocabulary. A Jaccard of 1.0 means identical vocabularies; 0.0 means no overlap. The 8×8 matrix reveals the linguistic topology of the corpus.

Findings

	CreAfE	ESol	MSIP	SigAfE	ANehr	ABurg	JCole	MDhs
CreAfE	100	42.6	28.5	25.9	33.9	8.8	23.3	7.8
ESol	42.6	100	28.8	30.3	70.6	10.3	23.1	8.7
MSIP	28.5	28.8	100	20.1	23.6	11.6	17.0	9.2
SigAfE	25.9	30.3	20.1	100	32.4	27.1	38.3	23.8
ANehr	33.9	70.6	23.6	32.4	100	12.1	26.3	11.2
ABurg	8.8	10.3	11.6	27.1	12.1	100	22.5	74.2
JCole	23.3	23.1	17.0	38.3	26.3	22.5	100	23.5
MDhs	7.8	8.7	9.2	23.8	11.2	74.2	23.5	100

TABLE F4-1: 8×8 Jaccard Vocabulary Similarity Matrix (%). Diagonal = 100% (self). Highest off-diagonal: ABurg ↔ MDhs (74.2%). Lowest: MDhs ↔ CreAfE (7.8%).

Top 5 Most Similar Group Pairs:

Rank	Group A	Group B	Jaccard %	Note
1	AUTH_Amanda-Burg	AUTH_Megan-Larison-DHS	74.2%	Both DHS-originated
2	AUTH_Alisha-Nehring	META_ESolutions-Sig	70.6%	Strongest META ↔ AUTH link
3	META_Created-after-eFile	META_ESolutions-Sig	42.6%	Largest META ↔ META pair
4	AUTH_Judith-Cole	META_Signed-after-eFile	38.3%	Cross-boundary
5	AUTH_Alisha-Nehring	META_Created-after-eFile	33.9%	Cross-boundary

META/AUTH Boundary Test:

Comparison	Avg Jaccard %	Interpretation
Within META (6 pairs)	29.4%	Moderate intra-class similarity
Within AUTH (6 pairs)	28.3%	Moderate intra-class similarity
Cross META ↔ AUTH (16 pairs)	23.1%	Lower cross-boundary similarity

Significance

The META/AUTH boundary is a modest language boundary: cross-class Jaccard (23.1%) is lower than within-class averages (29.4% META, 28.3% AUTH), but the difference is not dramatic. The strongest linguistic bond in the entire corpus is AUTH_Amanda-Burg ↔ AUTH_Megan-Larison-DHS (74.2%), both DHS-originated groups producing template-standardized competency evaluation documents. The second strongest is AUTH_Alisha-Nehring ↔ META_ESolutions-Signature (70.6%), crossing the META/AUTH boundary entirely — a connection driven by shared clinical/legal vocabulary in documents that happen to also carry Tyler/ESolutions digital signatures.

Connecting Context

Part 2's row-level Jaccard on hashpack_rows measured text-block sharing (exact content). The present string_norm Jaccard measures vocabulary sharing (normalized terms). That AUTH_Alisha-Nehring ↔ META_ESolutions-Signature shows high similarity at both levels (Part 2 cross-group hash overlap + Part 5 vocabulary overlap) strengthens the conclusion that these two groups share substantial textual content. See Appendix A5.

Finding 5: string_group_type Taxonomy Profiles

Background

The string_group_type classification assigns each extracted term to a semantic category. The distribution of categories per group reveals the dominant thematic content: clinical/diagnostic (dsm5), expert evaluation (experts), court authority (protect), commitment language (indefinite), or institutional logistics (itinerary).

Findings

Group	#1 SGT (%)	#2 SGT (%)	#3 SGT (%)
META_Created-after-eFile	experts (26.0%)	protect (16.9%)	truth (10.6%)
META_ESolutions-Sig	experts (18.5%)	dsm5 (16.6%)	protect (13.9%)
META_MSIP	protect (25.6%)	experts (12.6%)	dsm5 (9.9%)
META_Signed-after-eFile	experts (23.2%)	protect (22.9%)	itinerary (18.3%)
AUTH_Alisha-Nehring	dsm5 (19.7%)	experts (14.8%)	protect (10.5%)
AUTH_Amanda-Burg	itinerary (35.8%)	protect (20.0%)	experts (18.3%)
AUTH_Judith-Cole	experts (24.2%)	protect (21.2%)	indefinite (19.7%)
AUTH_Megan-Larison-DHS	itinerary (34.0%)	protect (20.4%)	experts (18.7%)

Significance

AUTH_Amanda-Burg and AUTH_Megan-Larison-DHS share the same top-3 taxonomy profile (itinerary, protect, experts) in nearly identical proportions, reinforcing their DHS template kinship. META_Created-after-eFile is the only group where 'truth' appears in the top 3 (10.6%), reflecting its unique vocabulary of defendant self-referential claims. AUTH_Alisha-Nehring is the only group led by 'dsm5' (19.7%), consistent with clinically detailed evaluation reports. 'fraud' appears in META_Created-after-eFile (5.9%) and META_MSIP (5.8%) but is negligible elsewhere.

Connecting Context

The taxonomy profiles confirm the classification logic: AUTH groups produce clinical/institutional documents (dsm5, itinerary, experts), while META groups — especially META_Created-after-eFile — capture a broader range including defendant-authored content (truth, fraud). See Appendix A6.

Finding 6: String Identity Trio Consistency

Background

The `string_norm` / `string_raw` / `string_id` trio enables verification that normalization is consistent and that identifiers map 1:1 to canonical forms. This analysis checks for normalization variance (multiple raw forms per norm), ID multiplicity (multiple IDs per norm), and ID collisions (one ID mapping to multiple norms).

Findings

Of the top 30 terms by occurrence, 25 have a strict 1:1 mapping across all three trio fields. Five terms show variation:

string_norm	Distinct Raw	Distinct ID	Note
Rule 20 AND Rule 20.01	3	1	3 raw variants normalized to 1 canonical form
MN Department of Human Services	3	1	Minor capitalization/spacing variants
credibility	4	1	4 raw surface forms, same ID
transport	2	1	Minor variant
disorder	1	2	Same norm, 2 IDs (dsm5 + control categories)
impairment	1	2	Same norm, 2 IDs (competency + dsm5 categories)

Zero `string_id` collisions: no `string_id` maps to more than one `string_norm` value. The multi-ID cases for 'disorder' and 'impairment' are by design: the `string_id` format encodes the `string_group_type`, so the same term classified under two semantic categories receives two structurally distinct IDs (e.g., `str:dsm5:sha256:4dba...` vs `str:control:sha256:4dba...`). The underlying SHA256 hash is identical in both cases.

Significance

The string identity trio is highly consistent. Raw-to-norm variance is minimal (5 of top 30 terms, all with 2–4 surface forms collapsing to 1 canonical form). No collisions or mapping anomalies were detected. This confirms the extraction pipeline produces reliable canonical identifiers suitable for cross-group comparison. See Appendix A7.

Finding 7: Row-Level Text Sharing Cross-Reference (Part 2 Bridge)

Background

Part 2 analyzed cross-group document sharing at the row-hash level (hashpack_rows). This finding bridges that analysis to the term level by examining what vocabulary (string_norm terms from children_doc_strings) appears in documents that share cross-group row hashes.

Findings

The connection between row-level hash sharing (Part 2) and term-level vocabulary overlap (Part 5) is structural: documents that share identical text blocks (row_sha256 matches) necessarily share the terminology extracted from those blocks. The Jaccard analysis in Finding 4 quantifies this at the vocabulary level. The highest cross-group Jaccard pairs (AUTH_Amanda-Burg ↔ AUTH_Megan-Larison-DHS at 74.2%; AUTH_Alisha-Nehring ↔ META_ESolutions-Signature at 70.6%) are consistent with the groups that showed the highest cross-group hash overlap in Part 2.

The 9 universal terms ('Rule 20 AND Rule 20.01', 'services', 'health', 'care', 'facility', 'mental health', 'participate', 'competency to proceed', 'By the court') form the linguistic backbone shared across all row-hash groups. These terms represent the minimum common vocabulary of the competency evaluation pipeline.

Significance

The term-level and hash-level analyses are convergent: groups with high vocabulary Jaccard also show high row-hash overlap. This dual confirmation strengthens the conclusion that cross-group content sharing is genuine textual overlap, not merely coincidental use of common terms. See Appendix A8.

Finding 8: source_url Analysis

Findings

100% of cohort string rows (33,645 rows across all 8 groups) carry a non-null string_source_url value. Every extracted term is linked to a specific page and text fragment in the source PDF document via a URL of the form: `https://mncourtfraud.com/file1/{filename}#page={N}&~:text={term}`.

These URLs use the browser Text Fragment specification to deep-link to the exact location where the term was extracted. This means every statistic in this report is independently verifiable by navigating to the source URL and confirming the term appears at the indicated page location.

Significance

The 100% source_url coverage provides an extraordinary level of provenance for extracted terminology. A third-party verifier can follow any string_source_url to visually confirm that the term appears in the cited document at the cited page. This is the strongest per-row provenance chain in the entire 5-part series. See Appendix A9.

Finding 9: Language Signature Cards Per Group

Background

Each group's language fingerprint synthesizes coverage, vocabulary size, exclusive terms, dominant taxonomy, top terms, cross-group similarity, and named persons into a single identity card.

Findings

META_Created-after-eFile

Metric	Value
Coverage	56.1%
Distinct Terms	291
Exclusive Terms	98 (33.7%)
Top Exclusive	Google, infiniset, compel discovery, synthetic, invention
Dominant SGT	experts (26.0%)
Top-3 Terms	services (809), Rule 20 (788), commitment (489)
Jaccard Max	META_ESolutions-Sig (42.6%)
Named Persons	Jill Rogstad (65 rows), Julia Dayton Klein (27), Lisa Janzen (27), Carolina Lamas (24), Thomas Arneson (22)
Linguistic Identity	The broadest and most linguistically distinctive group. Dominated by expert evaluation terminology but uniquely contains defendant self-referential vocabulary ('truth' category: Google, USPTO, CIA, LinkedIn) and system integrity terms ('fraud' category: synthetic, metadata, compel discovery) found in no other group.

META_ESolutions-Signature

Metric	Value
Coverage	98.5%
Distinct Terms	261
Exclusive Terms	19 (7.3%)
Top Exclusive	in the Interests of Justice, state-operated treatment, petition for judicial commitment, competency education coordinator, hypomania
Dominant SGT	experts (18.5%)
Top-3 Terms	Rule 20 (353), services (231), commitment (159)
Jaccard Max	AUTH_Alisha-Nehring (70.6%)
Named Persons	John Anderson (123), Carolina Lamas (40), Jill Rogstad (39), Julia Dayton Klein (38), Megan Paris (37)
Linguistic Identity	A terminologically diverse group with clinical/legal vocabulary nearly identical to AUTH_Alisha-Nehring (Jaccard 70.6%). Its 19 exclusive terms include specialized legal phrases ('in the Interests of Justice') and clinical terms ('hypomania', 'non-restorable', 'civil commitment is indefinite').

META_MSIP

Metric	Value
Coverage	100%
Distinct Terms	106
Exclusive Terms	8 (7.5%)
Top Exclusive	Rebecca Cole, Zachary Cole, cole v cole, controlling, MN Stat. 61142, constitutional violations, illogical, Elizabeth Murphy
Dominant SGT	protect (25.6%)
Top-3 Terms	duty (45), Rule 20 (45), By the court (32)
Jaccard Max	META_ESolutions-Sig (28.8%)
Named Persons	Emmett Donnelly (22), Katheryn Cranbrook (19), William Koch (10), Rebecca Cole (8), Raissa Carpenter (7)
Linguistic Identity	The appellate group: led by judicial authority language ('duty', 'By the court') with the highest attorney-name density. Its exclusive terms include unique person references (Rebecca Cole, Zachary Cole, cole v cole) and 'constitutional violations' — language specific to appellate review of constitutional claims.

META_Signed-after-eFile

Metric	Value
Coverage	95.0%
Distinct Terms	79
Exclusive Terms	0 (0.0%)
Top Exclusive	(none — entire vocabulary is a subset of other groups)
Dominant SGT	experts (23.2%)
Top-3 Terms	Rule 20 (39), MN DHS (33), services (33)
Jaccard Max	AUTH_Judith-Cole (38.3%)
Named Persons	Julia Dayton Klein (10), Carolina Lamas (8), Michael Browne (4), Sonia Reardon (2)
Linguistic Identity	A linguistically transparent group with zero exclusive terms — every term in its vocabulary appears in at least one other group. Its strongest vocabulary match is AUTH_Judith-Cole (38.3%), reflecting shared commitment/evaluation language in documents that were signed after their official e-file date.

AUTH_Alisha-Nehring

Metric	Value
Coverage	100%
Distinct Terms	203
Exclusive Terms	4 (2.0%)
Top Exclusive	behavioral health case manager, hallucination, non-compliance, Myles Antonioli
Dominant SGT	dsm5 (19.7%)
Top-3 Terms	Rule 20 (199), credibility (139), services (123)
Jaccard Max	META_ESolutions-Sig (70.6%)
Named Persons	John Anderson (117), Lisa Janzen (67), Stephanie Bruss (48), Megan Paris (36), Carolina Lamas (36)
Linguistic Identity	The most clinically detailed group, led by DSM-5 diagnostic terminology (19.7%). Its 70.6% vocabulary overlap with META_ESolutions-Signature and exclusive terms like 'hallucination' and 'non-compliance' confirm these are substantive forensic evaluation reports with extensive clinical content.

AUTH_Amanda-Burg

Metric	Value
Coverage	100%
Distinct Terms	29
Exclusive Terms	1 (3.4%)
Top Exclusive	Gregory Hanson (examiner, 12 rows in 3 docs)
Dominant SGT	itinerary (35.8%)
Top-3 Terms	MN DHS (976), Rule 20 (496), services (298)
Jaccard Max	AUTH_Megan-Larison-DHS (74.2%)
Named Persons	Lisa Janzen (36), Gregory Hanson (12)
Linguistic Identity	A highly template-standardized group with only 29 distinct terms, 74.2% vocabulary overlap with AUTH_Megan-Larison-DHS, and 26.8% of all rows devoted to a single term ('MN Department of Human Services'). Its sole exclusive term is the examiner name Gregory Hanson.

AUTH_Judith-Cole

Metric	Value
Coverage	100%
Distinct Terms	80
Exclusive Terms	4 (5.0%)
Top Exclusive	Kathryn Jameson, Chris Bowerman, Heidi Strohmaier, Jessica Miles (all examiners)
Dominant SGT	experts (24.2%)
Top-3 Terms	commitment (1,154), Rule 20 (880), services (825)
Jaccard Max	META_Signed-after-eFile (38.3%)
Named Persons	Lisa Janzen (214 rows, 105 docs), Sonia Reardon (68), Carolina Lamas (56), Kathryn Jameson (38)
Linguistic Identity	The largest terminology producer (10,655 rows) with Lisa Janzen appearing in 105 of 123 documents (85.4% doc penetration). Dominated by commitment and expert evaluation language. All 4 exclusive terms are examiner names, making examiner identity the sole differentiator.

AUTH_Megan-Larison-DHS

Metric	Value
Coverage	100%
Distinct Terms	25
Exclusive Terms	0 (0.0%)
Top Exclusive	(none)
Dominant SGT	itinerary (34.0%)
Top-3 Terms	MN DHS (189), Rule 20 (112), committed (58)
Jaccard Max	AUTH_Amanda-Burg (74.2%)
Named Persons	Lisa Janzen (12), Stephanie Bruss (6), Carolina Lamas (2)
Linguistic Identity	The smallest vocabulary in the corpus (25 terms, zero exclusive). A near-perfect vocabulary subset of AUTH_Amanda-Burg (74.2% Jaccard). DHS-originated template documents with standardized institutional language and no group-defining unique terminology.

Finding 10: Series Synthesis — Group Identity Summaries

Background

This is the capstone of the 5-part series. For each group, findings from all five parts are synthesized into a comprehensive identity description based exclusively on measured data from the live database.

Findings

META_Created-after-eFile (656 docs)

Dimension	Data
Part 1: Filing Profile	Largest group at 656 docs (62.9% of cohort). Spans 8 years (2017–2025) across 173 cases and 63 clusters. Broadest filing type coverage: top filing type (Notice of Case Reassignment) accounts for only 21.6%.
Part 2: Content Hashes	Highest proportion of unique row hashes. Content uniqueness is elevated relative to group size.
Part 3: Object Inventory	Contains TRACKING-FONT generations A through F. Most structurally diverse PDF object inventory.
Part 4: Signatures/Persons	99.7% system-only signatures (MCRO Watermark). Virtually no judicial officer signature presence (0.3%). Avg signing delay: 719 days.
Part 5: Language/Terms	291 distinct terms, 98 exclusive (33.7%). Dominated by 'experts' (26.0%) and uniquely contains 'truth' (10.6%) and 'fraud' (5.9%) categories. Exclusive vocabulary includes Google, infiniset, USPTO, CIA, LinkedIn, synthetic, metadata, compel discovery.
SYNTHESIS	META_Created-after-eFile is the broadest, most diverse, and most linguistically distinctive group in the corpus. It is defined by a metadata anomaly (PDF creation date post-dating e-file date), contains documents spanning the widest temporal and case range, produces the most unique content at both the hash and term level, and is the sole repository of defendant self-referential and system-integrity vocabulary. No other group shares any of its 98 exclusive terms.

META_ESolutions-Signature (134 docs)

Dimension	Data
Part 1: Filing Profile	134 docs across 76 cases and 47 clusters. Spans 2,759 days (2017–2025). Broadest cross-group overlap hub (appears in 5 of 7 non-zero off-diagonal cells in Part 1 co-occurrence matrix).
Part 2: Content Hashes	High cross-group hash sharing, particularly with AUTH_Alisha-Nehring and META_Signed-after-eFile.
Part 3: Object Inventory	Contains Tyler/ESolutions signature objects. RSA-1024 deprecated key detected in root MSIP document.
Part 4: Signatures/Persons	100% JO signature presence. 79.0% edits-after-signing. 77.1% NULL cryptographic validity. Avg 4.75 sigs/doc. Julia Dayton Klein: 31 docs (broadest JO).
Part 5: Language/Terms	261 terms, 19 exclusive (7.3%). Top Jaccard: AUTH_Alisha-Nehring (70.6%). Exclusive terms include legal phrases and clinical specialties (hypomania, non-restorable, civil commitment is indefinite).
SYNTHESIS	META_ESolutions-Signature is the cross-group connectivity hub of the corpus. It has the highest average signature count (4.75/doc), the highest edits-after-signing rate (79.0%), and the strongest vocabulary link to an AUTH group (70.6% Jaccard with AUTH_Alisha-Nehring). Its 19 exclusive terms are specialized legal and clinical phrases absent from template-based groups.

META_MSIP (34 docs)

Dimension	Data
Part 1: Filing Profile	Smallest META group. 79.4% MN Court of Appeals case type. 141-day temporal window (Apr–Aug 2025). Only group with significant null cluster_id (82% appellate filings).
Part 2: Content Hashes	Limited cross-group hash overlap due to small size and appellate document type.
Part 3: Object Inventory	Contains MSIP enterprise labels from a single M365 tenant spanning 3 judicial branches.
Part 4: Signatures/Persons	67.6% signature coverage (11 unsigned docs). 100% NULL cryptographic validity. Jennifer Frisch exclusive to this group.
Part 5: Language/Terms	106 terms, 8 exclusive (7.5%). Led by 'protect' (25.6%). Exclusive terms: Rebecca Cole, Zachary Cole, cole v cole, constitutional violations, controlling, illogical, Elizabeth Murphy, MN Stat. 61142.
SYNTHESIS	META_MSIP is the appellate outlier: the only group dominated by Court of Appeals filings, the only group with significant unsigned documents, the only group with 100% NULL signature validity, and the only group with exclusive person-name terms (Rebecca Cole, Zachary Cole). Its vocabulary reflects appellate judicial authority language rather than clinical evaluation terminology.

META_Signed-after-eFile (20 docs)

Dimension	Data
Part 1: Filing Profile	20 docs across 16 cases. 2,148-day span (2018–2024). 18-document overlap with META_ESolutions-Signature (largest off-diagonal cell).
Part 2: Content Hashes	High hash overlap with META_ESolutions-Signature (the 18 shared documents).
Part 3: Object Inventory	Multi-signature structure consistent with documents bearing both ESolutions CA and MCRO signatures.
Part 4: Signatures/Persons	100% JO presence. 73.3% edits-after. 64.0% NULL validity. Highest median rev_count (6).
Part 5: Language/Terms	79 terms, 0 exclusive. Entire vocabulary is a subset of other groups. Strongest Jaccard: AUTH_Judith-Cole (38.3%).
SYNTHESIS	META_Signed-after-eFile is the linguistically transparent group — zero exclusive terms, meaning it introduces no unique vocabulary to the corpus. Its identity is defined entirely by its metadata property (signing after e-file date) and its structural overlap with META_ESolutions-Signature. It is a forensic intersection group, not a vocabulary-distinct group.

AUTH_Alisha-Nehring (29 docs)

Dimension	Data
Part 1: Filing Profile	29 docs across 24 cases. Strong overlap with META_ESolutions-Signature (10 docs).
Part 2: Content Hashes	Elevated cross-group hash sharing with META_ESolutions-Signature.
Part 3: Object Inventory	Multi-signature objects consistent with ESolutions CA presence.
Part 4: Signatures/Persons	100% sig coverage. 83.7% NULL validity. 77.5% edits-after. Avg signing delay: 197 days (bimodal).
Part 5: Language/Terms	203 terms, 4 exclusive (hallucination, non-compliance, behavioral health case manager, Myles Antonioli). DSM-5 led (19.7%). Jaccard max: META_ESolutions-Sig (70.6%). Dense named-person content: John Anderson (117 rows), Lisa Janzen (67).
SYNTHESIS	AUTH_Alisha-Nehring is the most clinically detailed group in the corpus, led by DSM-5 diagnostic terminology and featuring the richest named-examiner vocabulary. Its 70.6% vocabulary overlap with META_ESolutions-Signature and exclusive clinical terms (hallucination, non-compliance) confirm these are substantive forensic evaluation reports with extensive clinical narrative content.

AUTH_Amanda-Burg (75 docs)

Dimension	Data
Part 1: Filing Profile	75 docs across 38 cases. 54.7% Proposed Order or Document filing type.
Part 2: Content Hashes	Template-level hash sharing consistent with standardized document production.
Part 3: Object Inventory	Standard PDF object structure without ESolutions multi-signature complexity.
Part 4: Signatures/Persons	100% sig. 64.8% valid. 30.6% edits-after. Avg delay: 351 days.
Part 5: Language/Terms	Only 29 distinct terms. 74.2% Jaccard with AUTH_Megan-Larison-DHS. Top term: MN DHS (976 rows, 26.8%). Sole exclusive: Gregory Hanson (examiner).
SYNTHESIS	AUTH_Amanda-Burg is a highly standardized DHS template group — only 29 distinct terms, 74.2% vocabulary overlap with AUTH_Megan-Larison-DHS, and over a quarter of all term occurrences devoted to a single institutional reference. Its sole linguistic differentiator from AUTH_Megan-Larison-DHS is the examiner name Gregory Hanson.

AUTH_Judith-Cole (123 docs)

Dimension	Data
Part 1: Filing Profile	123 docs across 86 cases. 85.4% single filing type (Findings and Order).
Part 2: Content Hashes	High intra-group hash consistency, moderate cross-group sharing.
Part 3: Object Inventory	Consistent object structure across the Findings and Order template.
Part 4: Signatures/Persons	100% sig. 58.9% valid. Lisa Janzen: 69 docs as JO signer.
Part 5: Language/Terms	80 terms, 4 exclusive (all examiner names: Kathryn Jameson, Chris Bowerman, Heidi Strohmaier, Jessica Miles). Lisa Janzen appears in 105 of 123 docs as a string_norm term (85.4%). Largest row producer: 10,655 rows.
SYNTHESIS	AUTH_Judith-Cole is defined by its procedural uniformity (85.4% single filing type) and the pervasive presence of Judge Lisa Janzen (85.4% document penetration as both JO signer in Part 4 and named term in Part 5). Its 4 exclusive terms are all examiner names, making professional identity the sole vocabulary differentiator. It produces more extracted terminology rows (10,655) than any other group despite having only 80 distinct terms.

AUTH_Megan-Larison-DHS (20 docs)

Dimension	Data
Part 1: Filing Profile	20 docs, 6 cases, 2 clusters. 508-day span. Zero overlap with any other group (most isolated).
Part 2: Content Hashes	Isolated hash universe consistent with zero document overlap.
Part 3: Object Inventory	Standard object structure. No ESolutions signatures.
Part 4: Signatures/Persons	100% valid. 0% edits-after. Avg signing delay: 1,383 days (~3.8 years). Strongest integrity profile.
Part 5: Language/Terms	25 terms, 0 exclusive. 74.2% Jaccard with AUTH_Amanda-Burg. Vocabulary is a near-complete subset of AUTH_Amanda-Burg.
SYNTHESIS	AUTH_Megan-Larison-DHS is the most isolated and most cryptographically clean group in the corpus: zero document overlap with any other group (Part 1), strongest signature integrity (Part 4), and zero exclusive vocabulary (Part 5). Its linguistic identity is indistinguishable from AUTH_Amanda-Burg (74.2% Jaccard), differing only by the absence of the examiner name Gregory Hanson. The group's distinctiveness lies entirely in its metadata properties (DHS authorship, extreme signing delay) rather than its textual content.

LIMITATIONS & ALTERNATIVE EXPLANATIONS

Data Coverage Gaps

META_Created-after-eFile has only 56.1% coverage in children_doc_strings. All vocabulary findings for this group are based on 368 of 656 documents. The 288 uncovered documents may contain additional terminology not captured in this analysis.

Extraction Pipeline Dependency

children_doc_strings reflects the output of a specific text extraction and classification pipeline. Terms present in the PDFs but not captured by this pipeline would be absent from the analysis. The classification into string_group_type categories is a product of the extraction system's taxonomy, not an inherent property of the documents. Independent re-extraction using different NLP tools could yield different term inventories.

Multi-Group Counting

48 documents belong to multiple groups. Terms from these documents are counted toward all applicable groups. This inflates apparent cross-group vocabulary overlap for the 27 META+AUTH overlap documents identified in Part 1.

Plausible Benign Explanations

The high vocabulary overlap between AUTH_Amanda-Burg and AUTH_Megan-Larison-DHS (74.2% Jaccard) likely reflects standardized DHS document templates used for competency evaluation reports. The 'truth' and 'fraud' exclusive vocabulary in META_Created-after-eFile may reflect defendant-authored motions and pro se filings that use non-standard legal terminology. The 98 exclusive terms in that group may simply reflect broader document type coverage rather than any systemic anomaly.

RECOMMENDATIONS FOR INDEPENDENT VERIFICATION

1. Independent NLP re-extraction: Re-extract terminology from the 1,043 cohort PDFs using an independent NLP pipeline (e.g., spaCy, Stanford NLP) and compare the resulting term inventories against `children_doc_strings`. Discrepancies would indicate extraction pipeline bias or coverage gaps.
2. META_Created-after-eFile coverage audit: Determine why 288 of 656 documents lack string data. Examine whether these documents have extractable text content or are image-only/scanned PDFs.
3. Exclusive term verification: For each of the 98 META_Created-after-eFile exclusive terms, manually verify their presence in the source PDFs via the `string_source_url` hyperlinks. Confirm they do not appear in documents from other groups that may have been excluded by pipeline coverage gaps.
4. DHS template analysis: Obtain the standard DHS document templates used for competency evaluation reports and compare their fixed text against the 25-term AUTH_Megan-Larison-DHS vocabulary. This would confirm whether the vocabulary is entirely template-derived.
5. Cross-reference named persons: Compare the examiner names appearing as `string_norm` terms (Part 5) against Part 4's judicial officer and examiner analysis. Verify that persons identified through text extraction match persons identified through signature metadata.
6. SQL reproducibility: All analyses are fully reproducible via the Anchor CTE and Appendix queries (A1–A10) against the live Supabase PostgreSQL database.

Possible Explanations (Unresolved; Not Proven by This Report)

- The 'truth' and 'fraud' terminology exclusive to META_Created-after-eFile may originate from defendant pro se filings or motions that reference personal claims and system integrity concerns. This is plausible but unverified.
- The near-identical vocabulary profiles of AUTH_Amanda-Burg and AUTH_Megan-Larison-DHS may reflect a single shared DHS template with minor authorship variation.
- The 70.6% vocabulary overlap between AUTH_Alisha-Nehring and META_ESolutions-Signature may reflect that documents with Tyler/ESolutions signatures are disproportionately clinical evaluation reports that happen to share Nehring's vocabulary profile.
- META_MSIP's exclusive 'person'-category terms (Rebecca Cole, Zachary Cole, cole v cole) may relate to a specific case involving these individuals that generated appellate filings captured by the MSIP metadata criterion.

APPENDIX | REPRODUCIBLE SQL QUERIES

All queries are SELECT-only, executed against Supabase project ibfmjtwahkwqzcmeyqii (PostgreSQL, us-west-2), March 2026. NOTE: children_doc_set column is “value” (not “doc_set”). children_doc_strings uses “string_source_url” (not “source_url”).

A1 — Anchor CTE (Cohort Definition)

```
WITH doc_set_cohort AS (  
  SELECT cds.evidence_uid, cds.value AS 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 — Coverage Analysis (Finding 1)

```
-- Per group: total docs, docs with strings, coverage %, rows, terms, SGTs  
-- LEFT JOIN doc_set_cohort to children_doc_strings on evidence_uid  
-- GROUP BY doc_set with conditional aggregation for coverage metrics  
-- Plus rows-per-doc: AVG and PERCENTILE_CONT(0.5) for median
```

A3 — Term Inventory & Concentration (Finding 2)

```
-- Per group: ROW_NUMBER() OVER (PARTITION BY doc_set ORDER BY COUNT(*) DESC) for top-  
20  
-- Concentration: SUM(top-10 rows) / total_rows per group  
-- GROUP BY doc_set, string_norm, string_group_type
```

A4 — Exclusive vs Shared Terms (Finding 3)

```
-- Step 1: DISTINCT doc_set, string_norm pairs  
-- Step 2: COUNT(DISTINCT doc_set) per string_norm -> group_count  
-- Step 3: Exclusive = group_count = 1; Universal = group_count = 8  
-- META-only = BOOL_OR(LIKE 'META_%') AND NOT BOOL_OR(LIKE 'AUTH_%')  
-- AUTH-only = inverse
```

A5 — Jaccard Vocabulary Similarity Matrix (Finding 4)

```
-- DISTINCT doc_set, string_norm pairs as term_groups  
-- Self-join: intersection = COUNT matching string_norms  
-- Union = size_A + size_B - intersection  
-- Jaccard = intersection / union * 100  
-- All 28 unique pairs (grp_a <= grp_b)
```

A6 — string_group_type Taxonomy Distribution (Finding 5)

```
-- Per group per string_group_type: COUNT(*) as row_count
-- Per group total: SUM for percentage calculation
-- ORDER BY doc_set, row_count DESC
```

A7 — String Identity Trio Consistency (Finding 6)

```
-- Top 50 string_norm by total occurrence
-- Per norm: COUNT(DISTINCT string_raw), COUNT(DISTINCT string_id)
-- Collision check: string_id -> COUNT(DISTINCT string_norm) HAVING > 1
-- Multi-ID check: string_norm with 2+ IDs -> show group_type split
```

A8 — Row-Level Text Sharing Cross-Reference (Finding 7)

```
-- Cross-reference approach: Join hashpack_rows to doc_set_cohort
-- Identify row_sha256 values appearing in documents from 2+ groups
-- Join back to children_doc_strings to find term overlap in shared-hash docs
```

A9 — source_url Analysis (Finding 8)

```
-- Per group: SUM(CASE WHEN string_source_url IS NOT NULL THEN 1 ELSE 0 END)
-- Top source_url values by frequency
-- All URLs in form: https://mncourtfraud.com/file1/{filename}#page={N}&::~text={term}
```

A10 — Named Persons in String Data (Finding 9)

```
-- WHERE string_group_type IN ('examiner', 'officer', 'attorney', 'person')
-- Per group: string_norm, string_group_type, COUNT(*), COUNT(DISTINCT evidence_uid)
-- Cross-reference with Part 4 judicial officer and examiner findings
```

End of Appendix. All queries verified against live database project ibfmjtwahkwqzcmeyqii on 2026-03-05.