

Methods companion

How the epoch-rollover screen of the ETSI corpus was built, validated,
and what its numbers do and do not support

Christian (fukami) Horchert

Data snapshot 2026-07-23 15:37

Licensed [CC BY 4.0](#)

Published at <https://lab.fukami.eu/epoch-scan/>

First published 2026-09-20 · revised 2026-09-21

What this document is. The companion to **Epoch rollover and the ETSI corpus**. That document is written for a technical committee deciding whether a work item is worth opening, and it deliberately carries no method detail. This one carries the method: how the corpus was cut, how the search vocabulary was built and revised, what a human read at a flagged clause consisted of, what the instrument was tested against, and where it is known to be blind. It assumes a reader who wants to check the instrument or reuse it, and it uses the project's own vocabulary rather than the paper's plainer register.

What the instrument is and is not

Stated plainly: every text-bearing deliverable was matched against a pattern registry of time-representation vocabulary (v3.2); each document was scored for *exposure*, meaning time-encoding vocabulary present, split by whether the document defines, inherits, or merely uses the encoding, and for *awareness*, meaning rollover discourse present; the documents that are exposed yet rollover-silent were flagged; the flagged set was ranked by native-defined exposure, because responsibility for a silent horizon lies with the document that defines the encoding; and the top of the ranking was read by hand at the flagged constructs.

The screen is a prioritisation instrument and not the fitness check itself. This distinction carries the whole method and is worth stating before anything else. Screening is automated and makes no safety claim about what it drops. Everything it produces is a candidate for reading, and everything below the reading layer in the paper passed a human gate at the clause. A reader who takes the ranking as a verdict has misread the instrument, and the funnel in §5 is drawn precisely so that the attrition is visible rather than implied.

Prior art, and why the new part is narrow

The IETF's Year 2000 Working Group ran essentially this method in 1999. RFC 2626 (*The Internet and the Millennium Problem*, Informational, Nesser) screened the RFC corpus for Y2K and period problems with a case-insensitive keyword filter, a Perl first cut given in its Appendix B, grouped the documents by service area, and put every hit through manual review. It stated outright that the hits "simply identify elements that need to be examined", and it flagged the signed-32 2038 class by name. The screen, the gate, and the candidate-versus-finding rule are therefore twenty-six years old. They were re-derived here from practice before RFC 2626 was known to the author, and that convergence reads as the design fitting the problem rather than as novelty.

The requirements-engineering literature owns the neighbouring ground. Lexical screening of specification text for quality defects is two decades deep, with NASA ARM, QuARS, and the requirements-smells line (Femmer et al., JSS 2017), and all of it detects defects that are *present* in the text. Missing requirements as a first-class analysis object belongs to van Lamsweerde's obstacle analysis, which works through formal goal

models rather than lexical screening. Detecting absence by corpus comparison is active current work: Luitel et al. (REFSQ 2023) use a masked language model to predict terminology missing from a requirement relative to a domain corpus, with a human in the loop.

What is new here is a narrow pair. The first is the *in-document structural precondition*: the absence flag is conditioned on the document’s own encoding definitions, so a specification that defines a 16-bit MJD field ($E_{\text{nat}} > 0$) and then says nothing about that field’s horizon is the subject. Trigger and expected mitigation live in the same document, and the cohort only calibrates rates. That precondition is what turns RFC 2626’s presence detector into a silence ranker, and it is not Luitel’s corpus-derived expected-vocabulary gap. The second is the *exposure-provenance split* ($E_{\text{nat}} / E_{\text{inh}} / E_{\text{use}}$), which assigns gap responsibility to the definer of an encoding rather than to its referencer or user, and for which no analogue was found in the completeness literature or the Y2K surveys. The honest word for the rest is recombination: RFC 2626’s screen-and-gate, the completeness literature’s absence-as-signal, and a provenance split that neither of them had.

Corpus and cohorts

Ownership	Deliverables	Scanned	Notes
etsi-native	1239	1238	named modern collections (not gate-read)
etsi-classic	16873	16055	legacy numbered trees (head gate-read + cheap-screened tail)
3gpp-transposed	49010	49009	TS/TR 120–149; gap head gate-read to completion
Total	67122	66302	66425 Published · 673 Historical · 24 On Approval

Cohort membership is structural and was fixed before any scoring. The 3gpp-transposed cohort is the TS/TR 120–149 transposition ranges, in which ETSI publishes 3GPP-authored text under an ETSI badge. The remainder splits into the named modern collections (etsi-native) and the legacy numbered trees (etsi-classic) along ETSI’s own numbering structure, with one known edge: the 100–109 GSM overlap is classed etsi-classic. Membership derives from document identity alone, so no cohort boundary was drawn after seeing any score. That ordering is the point of recording it here: a cohort fixed after scoring would make every per-cohort rate suspect.

The gap deep-dive targets the *ETSI-owned* deliverables, meaning the etsi-native and etsi-classic cohorts together. The 3GPP-transposed set is authored and maintained upstream at 3GPP, and ETSI carries only the badge, so its results are framed as a liaison contribution that points at where to look rather than as ETSI-owned defects.

Units of count

A *deliverable* is one published document version, and the *scanned* set is its text-bearing subset. A *family* is deliverable identity held across editions and across TS/TR/EN type conversions. *ETSI-owned* means the native and classic cohorts together, with the transposed cohort counted separately.

Family-level figures come in two named aggregations. Under *any-version*, a family qualifies if any one edition does, which means the figure includes families whose qualifying edition is not the newest. Under *latest-edition*, there is one row per family, its newest scored edition. Where the two units differ they are reported side by side (1129 against 1067 gap families). Family dedup matters because versions of one document are correlated observations, so per-deliverable counts inflate: a construct defined once and carried through eleven editions is one subject, not eleven.

In-force status

Publication status (Published, Historical, On Approval) is carried per deliverable from the catalogue and is reported alongside the counts above rather than folded into them. The status breakdown in that table describes the catalogue, not the scanned set, and the distinction matters: every scored document is Published, and not one Historical or On Approval deliverable carried extractable text into the scan. So the screen is silent about them by construction rather than by finding, which is a real gap, because a Historical edition can still describe deployed behaviour and an On Approval work item is the cheapest place for a horizon to be fixed. Near-publication flags come from the tracked work-item snapshot (`generated/active-wi-flags.csv`), not from the corpus store, because the portal state is a different source with a different refresh cadence.

The registry

A registry row carries an `id`, a `tier`, an `fm` tag, an `axis`, and a `regex`. `tier` is a precision prior, with 1 strict and 3 generic. `axis` routes a pattern into exposure (native, inherited or usage), into awareness, or into the cyclic reader column. `fm` tags the related XSTP.epoch failure-mode class and is a search-design aid, never an adjudication. The registry in force for every score reported here is v3.2, and it is reproduced verbatim from `generated/patterns.tsv` rather than retyped.

The registry was tuned on ETSI and broadcast text, so running the same instrument over the RFC corpus and a web-platform browser-specs corpus was a deliberate stress test of its vocabulary. The yield was ten IETF-mined and six web-mined pattern candidates, each adversarially reviewed, cross-checked against both corpora, and adjudicated with provenance and guard decisions. It flowed back as v3 (promoted 2026-07-14 after a blocking-gate verification pass) and v3.1 (three web candidates promoted after per-candidate adjudication, two deferred with tested guard designs, one dropped). The register-then-promote discipline held, since the shared instrument changed only at the calibration point. Three lessons are worth stating as lessons rather than as history.

Counter-wrap is not rollover awareness, and this was the expensive one. “Wraps around to zero” on a deliberately cyclic counter (RTP sequence numbers, the SRTP Roll-Over Counter, five-bit version counters) scored as awareness under v2. Every edition of EN 300 468 carries such text about table-version counters while staying silent on the date field, and TS 102 474’s entire awareness score was SRTP ROC vocabulary, which concealed a confirmed 16-bit MJD case in the review quadrant. Registry v3 moves this discourse to a cyclic axis scored into neither A nor E. It is separated rather than discarded: it stays visible to gate-readers as a per-document column, and deliberately-cyclic constructs remain separately reportable as a distinct class, one that is out of scope for *date-horizon* silence specifically. The fix is measurable in the live scores. EN 300 468 stands at A = 0 with CY = 18.4, so its residual wrap language is visible but no longer counts as rollover awareness, and TS 102 474 (CY = 47.1) sits in the gap quadrant it had been masked out of.

Vocabulary one culture makes explicit is silent in another. The 3GPP era idiom (“era 0”, “era 1”, “prime epoch”) and IETF-native constructs (TimeTicks, the NTP timestamp idiom, serial-number arithmetic) were invisible to the v2 patterns, while web-native time types have no ETSI footprint at all. The mined candidates were adjudicated and, where clean, promoted with a provenance assignment and context guards, on the principle that a duration is not an instant and a field named in days is not a time field. The ordinal-date form registered from the TS 129 274 miss moved that specification and its sibling out of the false-gap class.

Consequence for claims. Cross-corpus comparison of awareness scores is invalid. The registers differ, pattern weights do not transfer between corpora, and under-yield on a foreign corpus may mean missing vocabulary rather than absent exposure. This is why the paper reports the ITU-T awareness census as a statement about the search vocabulary’s reach as much as about that corpus.

The human gate

Every claim below the screening layer passes a manual gate: the specification text is read at the flagged construct before anything is called a finding. The protocol is fixed and was locked before reading began.

1. Locate every time or date encoding the document *defines* (fields, tables, ASN.1 or schema types, formats). Record width, units, epoch, and whether modulo or wrap behaviour is specified.
2. Search the document for boundary handling, including annexes: rollover, wrap, overflow, era, range, migration, extension mechanisms.
3. Check the references clause. If the encoding is imported, note the referenced document and whether the reference is dated or undated.
4. Check for a companion or test specification in the same family that could carry the handling.
5. Record the verdict with an evidence locator, meaning edition plus clause or table. **A verdict without a locator is invalid.**

The time budget is about fifteen minutes per document. If a document is unresolved at budget the verdict is unresolved with a note on what remains, and that outcome is data rather than failure. The verdict vocabulary is closed and was locked at the same time, distinguishing representation exhaustion from specified wrap, silence from safe-by-reference, and awareness that covers the exposed field from awareness that does not.

What “confirmed” means

Confirmed means exactly *confirmed at the flagged clause*. It is not a final verdict. Every finding in the paper is a pointer worth an expert look, and the actual verification step is expert review by the owning technical body. What the work hands over is a prioritised, evidence-located review queue and nothing more.

The readers are careful non-specialists. The gate is a clause-level read by readers without deep experience in the individual standards, not domain-expert adjudication. Non-obvious misses are expected, because a reader without standard-specific experience can miss context an expert would catch, such as a companion specification, a deployment practice, or an editorial convention. That limitation is disclosed rather than mitigated, and it sets the ceiling on every verdict in the record.

Sampling and what each stratum estimates

The manual sample was drawn with a fixed seed over the ETSI-owned subset, in six strata, with the family as the unit and its latest version as the representative. Sizes were a budget rather than a promise, read incrementally, with the head precision stratum and the near-threshold stratum first. Each stratum estimates something different and the estimates are never pooled: the gap-quadrant strata give per-stratum precision at family level; the below-threshold stratum tests for false negatives near the cut, where any confirmed case is direct evidence the cut loses real findings; the zero-exposure stratum probes the lexical blind spot, where any confirmed case bounds structural rather than threshold recall; and the review-quadrant stratum asks whether awareness language actually covers the exposed fields. No global precision or recall claim is made from these counts.

The funnel, both sides

The screen is a funnel, not a measurement. The attrition from corpus to reviewable head is the method working as designed, on a log scale. The first two bars count the whole corpus; the lower three count ETSI-owned families only, so the step between the second and third changes unit as well as size:



ETSI-owned families: exposed and rollover-silent		1129
... with native-defined time encoding		302
reviewable head ($E_{\text{nat}} \geq 20$)		139

Everything in those bars is screening output. The funnel also has a reference line, and it is an uncomfortable one: a naive keyword grep over the same corpus, using MJD | Modified Julian | NTP | GPS week | Unix time | time_t | epoch case-insensitively and unscored, returns **771** of the 7467 ETSI-owned latest-edition families. Candidate generation was never the scarce step. Ranking and attribution are, which is what the validation in §6 has to demonstrate rather than assert.

The 771 figure is one of the few numbers in this document entered by hand rather than bound to generated/. It was computed 2026-07-15 with a registry-independent grep list and is stated with that date because it cannot regenerate with the rest.

The gate side

Reading queues were drawn from the ranked heads and from targeted strata around them. Counts are per stratum, the queues overlap, and the final row dedups across them:

Stratum	Queued	Adjudicated	No text	Reader
native head (frozen v2 queue)	89	7	0	gate protocol read, in part
head boundary entrants	6	6	0	gate protocol read
NTP-boundary addendum	6	2	0	gate protocol read, overlaps boundary stratum
classic DVB/MJD family	23	23	0	gate protocol read
3GPP-transposed head	122	114	8	gate protocol read, to completion
D1 awareness-blocked stratum	14	14	0	agent read, verdicts CANDIDATE pending gate (1 promoted: TS 103 221-2)
stratified manual sample (precision check)	75	75	0	gate protocol read
DISTINCT families across the five gate queues (deduped)	223	142	8	overlaps removed, including families keyed in both the catalogue and the source-filename form

The reader column is the stratum's attribution, not a per-verdict ledger. Rows marked *gate protocol read* were read under the protocol in §4 by the careful non-specialist readers described there, and the one stratum read differently says so. Per-verdict attribution is partly unreconstructible, because execution self-documents and supervision does not, so the column states what the record supports rather than promising a ledger that does not exist. The paper's Appendix C carries the same limit in its own terms.

Every stratum was read under the gate protocol above, with one exception: the awareness-blocked stratum was agent-read with verdicts recorded as candidates pending the gate, and one was promoted after gate reading. Of the 223 distinct families queued, 142 carry a recorded verdict and 8 were queued with no extractable text to read; the remainder is unadjudicated screening output and not findings. The verdict tallies are read, not computed, which is the deliberate exception used throughout for gate outcomes.

The confirmed cases are scoped to the ETSI corpus. The IETF stream carries its own separate count and the two must not be added together, a distinction that had to be corrected in the paper on 2026-08-16 after the phrasing read as corpus-wide.

The funnel is also not the only discovery path. TS 102 474 surfaced off-funnel through a masked-awareness audit under registry v2, and the false-negative mode that exposed was measured and fixed at the v3 promotion (§3). An instrument whose only discovery path is its own ranking cannot find its own blind spots.

Instrument validation

An external review made the point that the instrument itself had not been evaluated. The screen was therefore run against a ladder of baselines over an identical corpus state. The framing has to come before any number: with only three confirmed positives inside the ladder’s ranking universe, counting the confirmed *local gaps* and not TS 103 311-3, whose verdict is a long-tail verification at about 2081 rather than a local gap although the package’s `gate_status` string-matches it as confirmed, there is no defensible precision or recall estimate, so this is a *disagreement analysis*, which asks where the methods rank the adjudicated families differently and why. It is not an evaluation, and calling it one would be the single easiest way to misuse this document.

Two of the confirmed cases cannot appear in the ladder by construction. The carried-record case sits in the transposed cohort, and the acknowledged-and-deferred case is awareness-blocked and is reached through the stratum read below rather than through the gap ranking. The ladder also compares methods on the *latest edition* of each of 7467 ETSI-owned families, whereas the headline figures aggregate any-version per family. Both units are named wherever they appear.

- **B0, naive grep:** the keyword list, membership only, no ranking.
- **B1, total exposure:** the gap quadrant ranked by E, no provenance split.
- **B2, raw native count:** gap families ranked by unweighted native-axis hit count, with no tier, zone or IDF weighting.
- **B3, the shipped instrument:** gap families ranked by E_{nat} , head at $E_{\text{nat}} \geq 20$.

Method	Set size	top 10	top 25	top 100	Known FPs in top 25
B0 grep	771	—	—	—	unranked
B1 total E	1 067	2	2	3	0
B2 raw native	284	3	3	3	0
B3 shipped	131 (head)	2	3	3	1 (rank 23)

The columns count the three confirmed local gaps retrieved within each cut. The known-FP column counts only the two adjudicated false-positive families, which is a lower bound on the set’s false-positive content and not a rate. These rows are transcribed from the tracked ladder record, and are among the deliberate read-not-computed exceptions to the regeneration rule. The record was first run on 2026-07-15 against registry v3.1 and re-run read-only against the shipped v3.2 scores on 2026-09-21; the set sizes and the known-FP rank moved, the retrieval columns did not.

What B0 says the screen is for. All three confirmed cases grep-match, so no confirmed finding was beyond the reach of a keyword search. What a 771-family grep queue lacks is order and provenance. The demonstrated value of the instrument on this corpus is ranking and attribution, not raw filtering. The naive list also misses a class the instrument handles: the dependency-deferred archetype matches no B0 keyword at all, because its time vocabulary lies outside the list. Grep and the instrument fail differently and neither subsumes the other.

B2 is about as good as B3, and the ontology is doing the work. Raw unweighted native counts already place all three confirmed cases in the top eight, while the shipped scoring places them at ranks 1, 3 and 13. Stated plainly: on the available positives the shipped weighting is *not shown* to improve on a raw native count, and the demonstrated contribution is the native, inherited and usage split itself. The tier, zone and

IDF weights are corpus-local scaffolding, which is consistent with pattern weights not transferring between corpora (§3).

One selection-bias caveat is load-bearing. The manual gate queue was read in B3 rank order, so the adjudication density along the B3 ranking is a reading-order artefact. It is never quoted as evidence that B3 orders better than B2, whose interleaved top entries are simply unread.

The excluded population, read by hand

Partitioning the 771 grep-positive families by what blocks them puts 84 in the head, 52 in the gap quadrant below the head threshold, 211 in gap with $E_{\text{nat}} = 0$, 410 in the low-exposure non-gap set, and 14 awareness-blocked. That last stratum is the cyclic-axis fix measured on an identical set: under registry v2 it held 55 families, and it drained to 14 at the v3 promotion.

The drain itself proves little, because it is entailed by the fix. Moving cyclic vocabulary off the awareness axis necessarily empties the stratum, so the number sizes the masking mode rather than validating the instrument. The load-bearing check is the by-hand read that followed: the residual 14 were construct-read, no second masked native gap recurred, and that is a real test of the instrument against its own excluded population.

That read also surfaced a finding class a silence screen cannot reach by construction. A document that acknowledges its horizon and defers the fix scores awareness, and so never enters the exposed-but-silent queue. The class turned out to be live, in a mandated lawful-interception wire format, and the case was simultaneously a native-axis false negative on four registry surface forms. Those four forms are recorded as candidates for a later registry revision and are *not* in the registry that produced the scores reported here. A screen that reads only its own gap quadrant systematically misses the “we know, we deferred” population, which is arguably the most actionable one, since maintenance intent already exists.

The $E_{\text{nat}} = 0$ strata, 631 grep-positive families across quadrants, are where a provenance-classification error would hide, invisible to downstream checks. They are sized here and deliberately left unread.

Calibration circularity

The registry was revised during this work, using the same documents that were read by hand. Its coverage is therefore calibrated against those reads rather than tested against an independent set. Aligning the instrument with the gated census is calibration to the same labels, not independent validation, and no out-of-sample check was run. This is the long form of the limitation the paper states in its own limits section, and it is the reason the next subsection exists.

The independent validation this section does not contain

Every check above is internal. The ladder and the cyclic-axis drain largely verify the instrument’s own arithmetic, and even the load-bearing stratum read is a by-hand audit rather than a prospective test.

The design that would validate the screen independently is a *temporal back-test*: freeze the registry on an edition predating a rollover-fixing amendment, check that the frozen screen ranks that edition exposed-and-silent, and check that awareness appears once the SDO’s own amendment lands. The subject must be selected from the SDO’s change record and never from our own adjudication, so that ground truth comes from real maintenance behaviour rather than from circular reasoning.

The test was designed and the corpus searched for a subject. There is none. The strongest candidate supplies a stable “pre”, with nineteen editions in the scanned corpus across three decades at constant exposure and zero awareness: two legacy ETS editions dated 1995 and 1997, then seventeen as EN 300 468, V1.3.1 to V1.19.1, the newest from 2025. Its fix is unpublished, so no “post” exists. Widening to every multi-edition family surfaced fourteen pairs in which rollover vocabulary first appears in a later edition. The eleven ETSI-owned pairs among them all add cyclic field-local wrap language and none adds date-horizon awareness; the strongest three, read by hand, are a PKI certificate-rollover permission, a coding-arithmetic “modulo 2”, and a designed-

cyclic section counter, none of which is an epoch fix. The only aware-axis transition in the whole corpus is a transposed GSM test specification, and there the awareness concerns a different field than its exposure.

Outside this corpus a first runnable cross-SDO subject was pre-checked. The X.509 and PKIX century-pivot chain supplies two public pre/post pairs, and the frozen registry sees neither fix: the awareness axis matches neither SDO's fix vocabulary, and one edition's own fix phrase scores as new native exposure. The foreign subject falsifies the frozen registry before it can validate the screen, which is the correct order for an independent check, and it turns registry extension by corpus insiders from an assumption into a measured precondition of the test.

A second designed-and-unrun check belongs here. Separating citation selection from inheritance in the reference-border numbers would take an ablation over the reference edges, comparing a single-lineage registry against the full one and counting which verified findings are lost. That comparison was specified and not run, which is why the paper states the border observation as a description of a citation-selected snapshot rather than as a test between the two explanations.

The instrument's trust rests on its characterised failure modes and the demonstrated by-hand miss, not on a prospective test. That is a weaker claim than a validated screen, and it is the accurate one.

Coverage limits and known misses

Silence is not a gap, and awareness is not coverage. A concern can be handled by reference or in a companion specification, and a document aware of one range construct can be silent on another, because document-level scores conflate fields. Two of the confirmed cases were masked exactly this way, one until the provenance split and one until the cyclic-axis check. The manual gate is therefore part of the method rather than an afterthought, and per-field scoring is not part of this instrument.

A wire encoding is not a stored encoding. A timestamp that wraps on the wire may be re-encoded on retention, and the reverse also happens, so a stored-record horizon claim requires the storage specification.

Some registry rows are census-checked but not gate-verified. Under the register-then-promote discipline a row can be registered on census evidence before a gate read confirms it. No confirmed finding rests on such a row, but the provenance-native ones do contribute to the E_{nat} screening counts and rankings, so the candidate lists inherit their unverified precision.

Coverage is depth-stratified and says nothing about the unread. The gate-read head, the cheap-screened slices and the unread tail are disclosed per cohort, and the screen cannot show that unflagged or unread documents are safe. Formal artefacts, meaning ASN.1 modules, XML schemas and bit-field tables, are invisible to a prose scan, and text extraction can lose content. Both external reviews named this. No measured extraction-loss instance is in hand: an earlier claimed one was retracted, because the text survived extraction and the miss was a registry pattern-shape gap, an ordinal date form the pattern did not expect. The false-negative rate is therefore likely underestimated. This scan did not audit extraction coverage.

The false-positive and false-negative profile inverts the deployed-scan profile. XSTP.epoch characterises deployed-substrate remote scanning as low false-positive and high false-negative. This screen has the opposite profile, because it measures a different object: specification text, not running systems. It is high false-positive, since lexical matches fire on non-temporal constants such as a bare maximum-value word used as a CRC seed or an identifier sentinel, on substring collisions where a type name sits inside a longer identifier, and on cyclic-counter vocabulary read as horizon awareness. Each of those is a burned candidate the gate had to overturn. The profiles differ because a specification cannot be exposed the way a deployed system can. The screen ranks where a bounded time constraint is normatively originated and left unacknowledged, which is a maintenance-locus signal rather than a susceptibility measurement.

Known imperfections, stated rather than fixed. The 100–109 GSM overlap is classed etsi-classic by convention. Version ordering is the larger one, and it is not rare: sorting editions by the source path string rather than by publication date picks an older edition for 468 of the 7 467 ETSI-owned families, because the GSM-era release tracks carry a higher version number on an earlier date. Every figure in these documents orders by publication date. The data package shipped on 2026-09-21 ordered by path string in its family table, which is why a reader reproducing from it saw 1 057 gap families at the latest-edition unit rather than 1 067; that build is superseded.

The per-pattern known misses, including the worked false-negative exemplar found during the ITU calibration pass and the ordinal-date form above, are carried in the machine-readable known-misses table generated alongside this document.

Evidence packets

Every confirmed case has an evidence packet, and the packet format is fixed so that a reviewer can check a claim without re-reading the whole specification. The format came out of an external review that asked what a finding would have to carry before an expert could act on it.

1. **Identity and maintenance locus.** Which document, which edition, and which body owns the fix.
2. **Exact clause or table.** The locator, not a description of where to look.
3. **Verbatim normative text.** A short excerpt rather than a paraphrase, so the reader can see the wording the verdict rests on.
4. **Width and order.** Field width, byte or bit order, and units.
5. **Exhaustion against specified wrap.** Whether the representation runs out or the text specifies wrapping behaviour. These are different defects with different fixes and conflating them is the most common way to overstate a finding.
6. **First unrepresentable date.** Computed, with the computation shown.
7. **Decoder requirements.** What an implementation must do that the text does not say.
8. **Cross-version persistence.** Whether the construct is stable across editions, which separates a live defect from a historical one.
9. **Severity and provenance.** The verdict, and the record trail behind it.

Seven packets exist, covering the ETSI-side confirmed cases, two RFC subjects reached through the reference border, a longevity note, and a deployment-leads record. They cite identifiers, versions, clause locators and short excerpts rather than full text, which is what makes them shareable when the underlying corpus is not.

Reproduction

Everything numeric regenerates from three artefacts: the pipeline code, the pattern registry, and the corpus. The corpus is *not redistributable*. ETSI deliverables are individually downloadable free of charge but carry redistribution restrictions, so the reproduction model is on-premises rather than bundled.

A verifier takes an existing authorised local corpus, checks it against a published manifest with content hashes, re-runs the scan, and checks the evidence packets. That is also what makes the workflow reusable by another SDO on a corpus this project could not lawfully mirror: the instrument travels, the corpus does not.

What the data package carries

The package holds the registry at the version that produced the scores, the per-corpus active-pattern lists, per-document manifests with character counts per text zone and the SHA-256 of the extracted text, the raw hit matrices before weighting, the per-document and per-family score tables, and the scoring SQL itself. It

carries no document text. A third party with its own copy of a document can confirm it scans the same text by comparing hashes, without any text being redistributed.

It is licensed CC BY 4.0. It is not a reproducibility bundle in the strong sense: corpus texts, pipeline code and per-candidate evidence records are not in it.

Responsible replication

Reproduction does not require, and should not encourage, repeated uncontrolled acquisition from ETSI or other community-operated services. The preferred path is hash verification against a corpus the verifier already holds, so the acquisition stage defaults to a non-fetching verification mode. Any fresh acquisition should use conservative rate limits, caching and resumability, conditional requests where supported, an identifiable user agent, and coordination with the service operator for large runs.

Determinism, and the deliberate exceptions

The registry version and the thresholds are pinned, and every screening figure in this document and in the paper is regenerated from the analysis store rather than typed. There are deliberate exceptions, and they are exceptions rather than oversights: the gate verdicts, the ladder rows, the stratum partitions and the cross-corpus queue sizes are *read, not computed*, transcribed from tracked validation records, each traceable to a record in the working set. The naive-grep reference line in §5 is the one figure carrying its own computation date for the same reason. Where a number is read rather than generated, this document says so at the point of use.