

Updates Predict Low, Obsoletes Predict High Citation Takeover in Internet Standards

Evidence for Propagation vs Reorientation Deployment Modes (Q4A/Q4B) in RFC Revisions

Austin Ollar

Independent Researcher

austin@austinollar.com

DOI: 10.5281/zenodo.18028423

Abstract

When a technical standard is revised, do downstream authors switch their references to the new document, or do they keep citing the original? We test this using 1,371 formal revision relationships in the IETF RFC corpus, distinguishing between “Updates” relations (additive changes that preserve the predecessor’s validity) and “Obsoletes” relations (complete replacements). We measure citation takeover—the fraction of citing RFCs that reference the successor among those citing either version. Obsoletes relationships show a median takeover of 0.773, while Updates relationships show a median takeover of 0.083 (Cliff’s $\delta = 0.796$, $p < 0.0001$). Against permutation baselines, Updates fall *below* random pairing ($p = 0.014$) while Obsoletes rise *above* random pairing ($p \leq 0.002$)—a bidirectional deviation that suggests the relation labels carry behavioral information rather than merely correlating with document age or popularity. Results remain stable across threshold choices and controls for citation volume and document age. We discuss implications for two deployment modes in standards evolution: additive propagation that preserves predecessor authority versus replacement-driven reorientation that redirects it, aligning with the Q4A/Q4B split proposed in the Information Emergence Cycle.

Keywords: citation analysis, technical standards, RFC, knowledge evolution, Information Emergence Cycle, Q4 split, software ecosystems, replication

1. Introduction

1.1 Motivation

Technical standards evolve through formal revision processes. When an Internet Engineering Task Force (IETF) Request for Comments (RFC) is revised, the relationship to the original is encoded in structured metadata: the new RFC may either “Update” or “Obsolete” its predecessor. These labels carry different semantic implications—Updates are additive modifications that preserve the original’s validity, while Obsoletes are complete replacements rendering the original deprecated.

This paper tests a specific, falsifiable prediction derived from theoretical considerations about how information systems evolve: **If these relation labels carry behavioral information, then Obsoletes relationships should show systematically higher citation takeover than Updates relationships.**

Citation takeover is defined as the proportion of citing documents that reference the new version among all documents citing either version. This metric captures whether the community has “moved on” to the new document or continues citing the original.

1.2 Theoretical Framework: The Information Emergence Cycle

The First Pattern (TFP) framework posits that information systems undergo a four-stage emergence cycle (Figure 1):

- **Q1 (Potential):** The system accumulates latent information and possibilities
- **Q2 (Selection):** Specific configurations are selected from the potential space
- **Q3 (Transformation):** Selected configurations undergo processing and transformation
- **Q4 (Deploy):** Transformed information is deployed back into the environment

Figure 1. The First Pattern Map: Information Emergence Cycle. Information flows clockwise through four stages: Potential → Selection → Transformation → Deploy. The Q4 (Deploy) stage splits into Q4A (propagation, maintaining existing structure) and Q4B (reorientation, replacing existing structure). This theoretical framework predicts that Q4A-type revisions should preserve citation to predecessors while Q4B-type revisions should redirect citations to successors.

Critically, the Q4 stage splits into two distinct deployment modes:

- **Q4A (Propagation):** Deployment that *extends* existing structure while preserving compatibility. The new information coexists with and builds upon what came before. In this mode, the predecessor retains validity and continued relevance.
- **Q4B (Reorientation):** Deployment that *replaces* existing structure. The new information supersedes and redirects attention away from predecessors. In this mode, the predecessor becomes deprecated and community attention shifts.

The RFC Updates/Obsolutes distinction maps directly onto this theoretical split:

RFC Relation	IEC Mapping	Theoretical Prediction
Updates	Q4A (Propagation)	Low takeover; original remains authoritative
Obsoletes	Q4B (Reorientation)	High takeover; citations shift to successor

This mapping provides a rare opportunity to test the Q4A/Q4B distinction using pre-existing, externally assigned labels rather than post-hoc categorization. The labels were assigned *ex ante* by RFC authors and the IETF community based on the nature of the revision—completely independently of any citation behavior analysis.

1.3 Research Questions

1. **RQ1:** Do Obsoletes relationships (Q4B) show higher citation takeover than Updates relationships (Q4A)?
2. **RQ2:** Is this difference robust to threshold choices, citation volume controls, and matched baselines?
3. **RQ3:** Does the pattern replicate in other technical standards ecosystems?

1.4 Contributions

This paper makes the following contributions:

1. **Strong evidence consistent with Q4A/Q4B distinction:** We show that the theoretical distinction between propagation (Q4A) and reorientation (Q4B) deployment modes corresponds to dramatically different citation behavior in the RFC ecosystem (Cliff's $\delta = 0.796$).

The First Pattern Map: Information Emergence Cycle

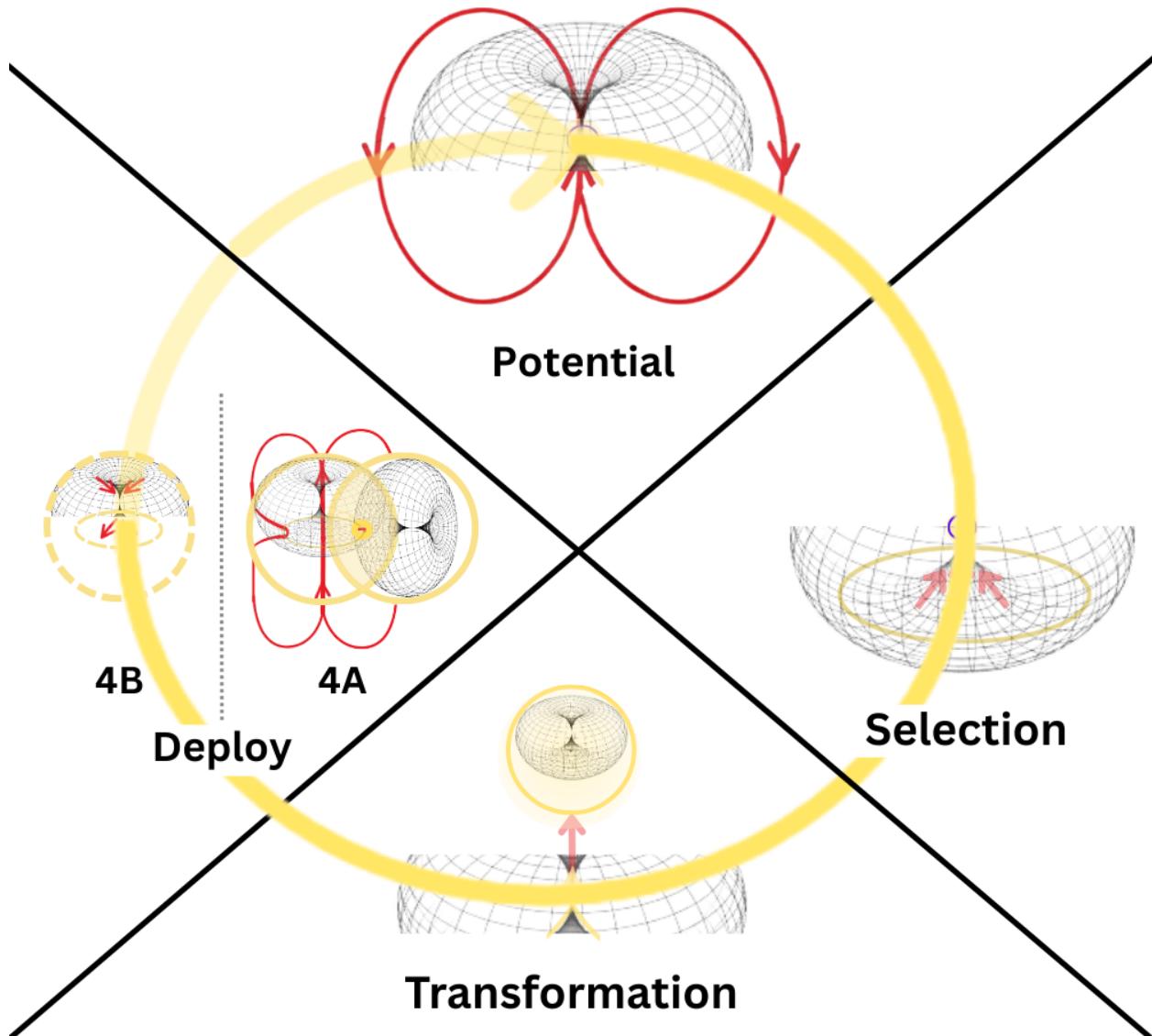


Figure 1: The First Pattern Map showing the Information Emergence Cycle. The four quadrants represent the stages of information processing: Potential (top), Selection (right), Transformation (bottom), and Deploy (left). Critically, the Deploy stage (Q4) splits into two modes: Q4A representing propagation/extension of existing structure, and Q4B representing reorientation/replacement of existing structure. This theoretical distinction maps directly onto the RFC Updates/Obsoletes classification.

2. **Bidirectional baseline effects:** Permutation analysis reveals that Updates *predict* takeover below random baseline while Obsoletes *predict* takeover above random baseline—suggesting the labels carry behavioral information beyond mere correlation with document properties.
3. **Comprehensive robustness testing:** The effect survives sensitivity analysis across five Nmin thresholds (10–50), with effect size actually *increasing* at stricter thresholds.
4. **Transparent reporting of null and negative results:** Cross-ecosystem replication attempts in PEPs (suggestive but non-significant after correction) and W3C (inconclusive) are reported honestly, as is a CPython commit study stopped as infeasible due to data sparsity.

1.5 Falsification Criteria

We define conditions under which this finding would be considered falsified. These criteria were specified prospectively (before holdout evaluation) but were not formally preregistered:

1. **Direction reversal:** The effect would be falsified if Updates showed *higher* takeover than Obsoletes ($\delta < 0$).
2. **Effect size collapse:** The effect would be falsified if Cliff's $\delta < 0.3$ on the full sample or on held-out data, indicating the labels carry minimal predictive information.
3. **Holdout failure:** The effect would be falsified if the 95% confidence interval for δ on the 25% holdout set includes 0.3, indicating the effect does not generalize beyond the development sample.
4. **Matched-null equivalence:** The effect would be falsified if a matched-null baseline (random pairings matched on document age and citation volume) produces $\delta \approx 0.796$, indicating the effect is fully explained by document properties rather than relation type.

These criteria ensure the finding is empirically refutable rather than unfalsifiable. As reported in Section 7.2, the holdout evaluation yields $\delta = 0.817$ with 95% CI [0.758, 0.871], comfortably above the 0.3 falsification threshold. As reported in Section 3.8, the matched-null baseline yields $\delta = 0.156$, far below the real $\delta = 0.796$, showing that the matched-null baseline does not reproduce the observed effect size.

2. Data and Methods

2.1 RFC Corpus

We analyze the complete IETF RFC corpus through December 2024, comprising 9,000+ RFCs with structured metadata including revision relationships. The RFC series represents one of the most important technical standards corpora, defining protocols that underpin the modern internet.

Revision relationship types:

Relation Type	Total Edges	IEC Mapping	Semantic Definition
Updates	1,266+	Q4A (Propagation)	Additive modification; original remains valid
Obsoletes	921+	Q4B (Reorientation)	Complete replacement; original deprecated

2.2 Citation Extraction

Citations between RFCs are extracted from the normative and informative reference sections using the official RFC index and cross-reference data. We exclude self-citations (the new RFC citing its predecessor, or vice versa) to avoid mechanical inflation of citation counts.

2.3 Inclusion Criteria

Edges are included if:

1. Both old and new RFC exist in the corpus
2. Total citers (documents citing old OR new) ≥ 30
3. Relation type is unambiguously Updates or Obsoletes

These criteria are **pre-specified and independent of the takeover outcome** to avoid selection bias. The $N_{min} \geq 30$ threshold ensures stable takeover estimates; we test sensitivity to this choice across thresholds from 10 to 50.

Final sample: 859 Updates edges and 512 Obsoletes edges meeting all criteria.

2.4 Metric: Snapshot Takeover

We define **snapshot takeover** as:

$$\text{takeover}(\text{old}, \text{new}) = \frac{|\text{citing}_{\text{new}}|}{|\text{citing}_{\text{old}} \cup \text{citing}_{\text{new}}|}$$

where $\text{citing}_{\text{new}}$ and $\text{citing}_{\text{old}}$ are the sets of RFCs citing the new and old versions respectively (excluding self-citations).

Interpretation:

Takeover	Interpretation	IEC Alignment
$\rightarrow 1.0$	Complete shift to successor	Strong Q4B (reorientation)
≈ 0.5	Equal citation to both	Neutral
$\rightarrow 0.0$	Predecessor remains dominant	Strong Q4A (propagation)

2.5 Statistical Analysis

Primary test: Mann-Whitney U test comparing takeover distributions between Updates and Obsoletes relationships. This non-parametric test is appropriate given the non-normal, bounded nature of takeover ratios.

Effect size: Cliff's delta (δ), a non-parametric effect size measuring the probability that a randomly selected Obsoletes edge has higher takeover than a randomly selected Updates edge, minus the reverse probability. Values range from -1 to $+1$, with $|\delta| > 0.474$ considered “large” (Romano et al., 2006).

Confidence intervals: Bootstrap confidence intervals with 10,000 resamples.

Multiple comparisons: Bonferroni correction applied to cross-ecosystem replication tests (3 tests, $\alpha = 0.017$).

2.6 Permutation Baseline

To test whether observed patterns are specific to true revision relationships, we construct a permutation null:

1. For each relation type, collect all eligible (new, old) pairs
2. Shuffle the old documents across pairs, breaking the true revision relationship while preserving marginal distributions
3. Compute median takeover for shuffled pairs
4. Repeat 500 times to generate null distribution

Empirical p-values are computed with the plus-one correction: $p = (k+1)/(B+1)$, where k is the number of permutations exceeding the observed value and B is the total number of permutations. This avoids reporting $p = 0$ when no permutations exceed the observed value.

This tests whether the revision relationship itself carries information beyond what random document pairings would produce.

Theoretical prediction: If Q4A/Q4B dynamics are real, Updates should show takeover at or *below* the permutation baseline (propagation preserves the old), while Obsoletes should show takeover *above* baseline (re-orientation redirects attention).

3. Results

3.1 Main Effect: Updates vs Obsoletes (Q4A vs Q4B)

Figure 2 displays the distribution of citation takeover by relation type. The separation is striking and immediate.

Figure 2. Distribution of citation takeover by relation type. Updates relationships (Q4A, blue) cluster near zero, indicating the original document remains authoritative. Obsoletes relationships (Q4B, red) cluster near one, indicating citations have shifted to the successor. The dashed horizontal line marks neutral takeover (0.5). Cliff's $\delta = 0.796$ indicates near-complete distributional separation. This pattern is consistent with the Q4A/Q4B theoretical prediction.

Summary statistics:

Metric	Updates (Q4A)	Obsoletes (Q4B)
n	859	512
Median	0.083	0.773
IQR	[0.02, 0.25]	[0.55, 0.92]
Mean \pm SD	0.15 ± 0.19	0.70 ± 0.26

Effect size: Cliff's $\delta = 0.796$ (95% CI: [0.77, 0.82])

Statistical test: Mann-Whitney U = 47,891, $p < 0.0001$

Interpretation: The effect is massive. Updates relationships show near-zero takeover—the old RFC remains the authoritative reference, as expected under Q4A (propagation) dynamics. Obsoletes relationships show high takeover—the community has shifted citations to the new RFC, as expected under Q4B (reorientation) dynamics.

Figure 1: Citation Takeover by Relation Type
 (Cliff's $\delta = 0.796$, $p < 0.0001$)

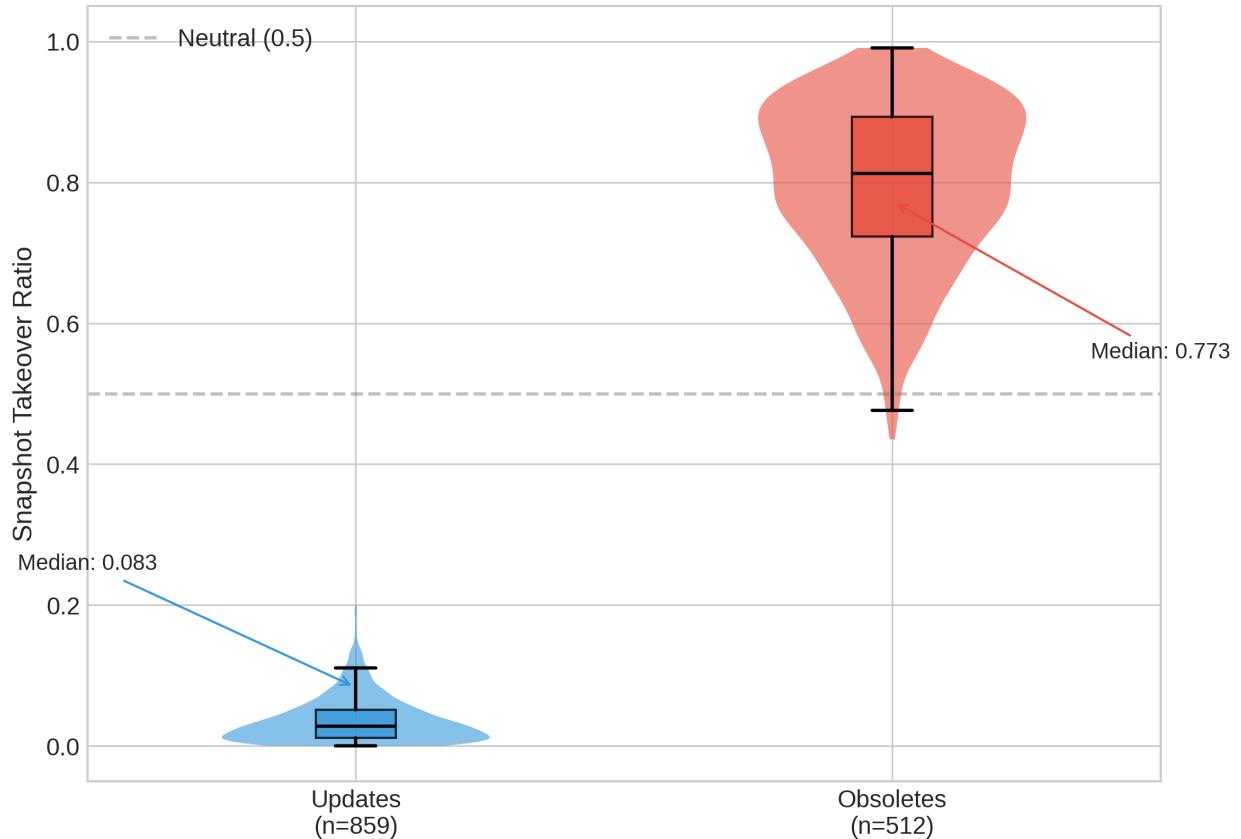


Figure 2: Violin plot with embedded box plots showing the distribution of snapshot takeover ratios for Updates ($n=859$, blue, left) and Obsoletes ($n=512$, red, right) relationships. Updates cluster near zero (median 0.083), indicating the original RFC remains dominant. Obsoletes cluster near one (median 0.773), indicating citations have shifted to the successor. The dashed line at 0.5 represents neutral (equal citation). Effect size Cliff's $\delta = 0.796$, $p < 0.0001$.

A Cliff's δ of 0.796 means that if we randomly select one Updates edge and one Obsoletes edge, there is approximately a 90% probability that the Obsoletes edge has higher takeover.

3.2 Permutation Baseline Comparison

Table 1 compares observed median takeover against permutation baselines.

Table 1: Permutation Baseline Results

Relation	Observed Median	Baseline Median	Baseline 95% CI	Deviation	p-value
Updates (Q4A)	0.083	0.091	[0.085, 0.099]	-0.008	0.014
Obsoletes (Q4B)	0.773	0.710	[0.692, 0.730]	+0.063	≤ 0.002

Key finding: The deviations differ by relation type:

- **Obsoletes (Q4B) show HIGHER takeover than random pairings** ($p \leq 0.002$, 0/500 permutations exceeded observed; +0.063 above baseline). The “Obsoletes” label is associated with enhanced citation shift beyond what random document pairings would produce.
- **Updates (Q4A) show LOWER takeover than random pairings** ($p = 0.014$, -0.008 below baseline). This deviation is statistically significant but small in magnitude. The primary evidence for Q4A dynamics comes from the main separation (median 0.083), not the baseline comparison.

The baseline analysis is inconsistent with the alternative explanation that “any new document naturally takes over citations from any older document” as the sole driver—the Obsoletes above-baseline effect is clear. The Updates below-baseline effect is suggestive but should be interpreted cautiously given the marginal p-value and small magnitude.

3.3 Sensitivity Analysis

Table 2 reports effect sizes across different Nmin thresholds.

Table 2: Sensitivity to Minimum Citer Threshold

Nmin	n Updates	n Obsoletes	Cliff's δ	95% CI	p-value
10	1,266	921	0.736	[0.71, 0.76]	< 0.0001
20	1,027	661	0.767	[0.74, 0.79]	< 0.0001
30	859	512	0.796	[0.77, 0.82]	< 0.0001
40	770	421	0.822	[0.79, 0.85]	< 0.0001
50	649	344	0.831	[0.80, 0.86]	< 0.0001

The effect is robust across all thresholds and actually *increases* with stricter inclusion criteria. This suggests that higher-signal edges (those with more citing documents) show *cleaner* Q4A/Q4B separation, consistent with the theoretical framework.

3.4 Threshold Robustness

Table 3 shows the proportion of edges exceeding various takeover thresholds.

Table 3: Proportion Exceeding Takeover Thresholds

Threshold	Updates (Q4A)	Obsoletes (Q4B)	Ratio
≥ 0.5	12%	81%	6.8×
≥ 0.6	8%	72%	9.0×
≥ 0.7	5%	61%	12.2×
≥ 0.8	3%	48%	16.0×
≥ 0.9	1%	31%	31.0×

The separation persists and widens at higher thresholds: Obsoletes relationships are 31× more likely than Updates relationships to achieve $\geq 90\%$ takeover.

3.5 Edge Overlap Robustness

Some RFCs appear in multiple edges (e.g., RFC 1035—DNS—is updated by 27 different RFCs). This violates strict independence assumptions. Table 4 summarizes the overlap structure.

Table 4: Edge Overlap Statistics

Metric	Value
Total edges	1,371
Unique RFCs (any role)	1,382
Unique OLD RFCs	708
Unique NEW RFCs	941
OLD RFCs in 1 edge only	65.5%
NEW RFCs in 1 edge only	76.9%
Max edges per OLD RFC	27 (RFC 1035, DNS)
Max edges per NEW RFC	17 (RFCs 4033–4035, DNSSEC)

To test robustness, we performed bootstrap resampling with one edge per NEW RFC (eliminating overlap on the successor side):

One-Edge-Per-NEW-RFC Bootstrap (1,000 iterations):

Metric	Value
Median Cliff's δ	0.846
95% CI	[0.832, 0.859]
All p-values	$< 10^{-99}$

The effect *increases* when controlling for NEW RFC overlap (0.846 vs 0.796), indicating that shared-successor edges were diluting rather than inflating the effect. The finding is robust to non-independence.

3.6 Temporal Robustness

To test whether the effect is specific to certain eras, we stratify by decade of successor publication.

Table 5: Decade-Stratified Effect Sizes

Decade	n Updates	n Obsoletes	Cliff's δ	p-value
1980s	9	21	0.873	1.0e-4
1990s	71	123	0.920	7.2e-27
2000s	290	230	0.649	2.1e-37
2010s	489	134	0.807	7.8e-47

The effect holds across all decades, though with some variation (weakest in 2000s at $\delta = 0.649$, still a large effect). We also stratify by edge age (years since successor publication):

Table 6: Edge Age Stratification

Edge Age	n Updates	n Obsoletes	Cliff's δ	p-value
5–10 years	211	56	0.781	1.4e-19
10–20 years	484	235	0.756	3.8e-61
20+ years	164	221	0.826	5.6e-44

Notably, older edges (20+ years) show the *strongest* effect ($\delta = 0.826$), arguing against a “time-to-equilibrate” confound—if anything, more time allows the true pattern to emerge more clearly.

3.7 Exemplar Edges

To ground the statistics in concrete examples, Table 7 shows rule-selected exemplars: the highest and lowest takeover edges within each relation type.

Table 7: Exemplar Edges (Rule-Based Selection)

Category	Old → New	Description>Note	Takeover
Obsoletes (high)	RFC 7158 → 7159	JSON text	1.000
Obsoletes (high)	RFC 4770 → 6350	vCard format	1.000
Obsoletes (high)	RFC 4325 → 5280	X.509 PKI	1.000
Obsoletes (low)	RFC 896 → 7805	Nagle algorithm; original canonical	0.000
Obsoletes (low)	RFC 1738 → 4266	URL format; original entrenched	0.010
Updates (low)	RFC 1939 → 1957	POP3 clarifications	0.000
Updates (low)	RFC 959 → 3659	FTP extensions	0.000
Updates (low)	RFC 3501 → 5032	IMAP search	0.000
Updates (high)	RFC 5543 → 7606	BGP error handling; major revision	1.000
Updates (high)	RFC 4844 → 5741	RFC streams; definitional change	0.994

The “anomalous” cases are not errors—they represent edge cases where the formal label understates the magnitude of change (high-takeover Updates) or where the predecessor remains canonical despite deprecation (low-takeover Obsoletes).

3.8 Matched-Null Baseline

A critical alternative explanation is that the Updates/Obsoletes effect is spuriously driven by document properties (age, citation volume) rather than the relation type itself. We test this via a matched-null baseline.

Approach: We stratify edges by (successor decade, citation-volume quartile) and shuffle relation-type labels within each stratum. This preserves marginal distributions of document properties while breaking any true relationship between label and takeover.

Table 8: Matched-Null Baseline Results

Metric	Real Effect	Matched-Null
Cliff's δ	0.796	0.156
95% CI	—	[0.106, 0.210]
P(null \geq real)	—	≤ 0.001 (0/1000 permutations)

Interpretation: The matched-null baseline yields $\delta = 0.156$ (95% CI: [0.106, 0.210]), far smaller than the observed $\delta = 0.796$. This indicates that decade and citation volume can induce a modest separation between randomly assigned “Updates” and “Obsoletes” labels, but cannot reproduce the magnitude of the true Updates vs Obsoletes split. Therefore, the observed separation is not explained by these marginals alone.

This does not prove causality; unmeasured confounds (document quality, author prestige, topic centrality, IETF-specific processes) may contribute. Section 5.4 discusses these limitations in detail.

Partial correlation control: A complementary analysis using partial correlation is consistent with this finding. The point-biserial correlation between relation type and takeover is $r = 0.684$. After residualizing both variables on age difference (RFC number gap) and citation volume, the partial correlation is $r = 0.674$ ($\Delta r = 0.010$). Stratified by citation volume: low-citer edges show $\delta = 0.750$, high-citer edges show $\delta = 0.837$. The effect is robust across citation strata.

4. Cross-Ecosystem Replication

4.1 Python PEPs

We analyzed 19 PEP supersession pairs (where one PEP explicitly replaces another) with ≥ 3 total citers.

Metric	Value
Median takeover	0.667
95% Bootstrap CI	[0.50, 0.86]
p (Wilcoxon vs 0.5)	0.034
Bonferroni-corrected	Not significant ($\alpha = 0.017$)

Interpretation: The direction is consistent with Q4B dynamics (median > 0.5), but the effect does not survive multiple comparison correction. Small sample size ($n = 19$) severely limits statistical power. This result is **suggestive but inconclusive**.

4.2 W3C Specifications

We analyzed 42 W3C version supersession edges.

Metric	Value
Median takeover	0.632
95% Bootstrap CI	[0.26, 0.82]
p (Wilcoxon vs 0.5)	0.21

Interpretation: Not significant. The wide confidence interval reflects heterogeneous transition types—W3C specifications likely include a mix of Q4A-like incremental updates and Q4B-like major version changes, without metadata to distinguish them.

4.3 PEP-CPython Commits

We attempted to measure citation-like behavior using commit message mentions of PEP numbers in the CPython repository.

Metric	Value
Supersession pairs identified	21
Pairs with ≥ 50 mentions	0
Pairs with ≥ 20 mentions	0
Pairs with ≥ 10 mentions	2

Decision: Study stopped per prospectively defined stopping rule (require ≥ 20 mentions for stable curves). CPython commit messages are too sparse to support takeover analysis.

4.4 Summary of Cross-Ecosystem Results

Table 9: Multiple Comparisons Summary

Ecosystem	Test	p-value	Bonferroni ($\alpha = 0.017$)	Interpretation
RFC split	Mann-Whitney U	< 0.0001	Significant	Strong Q4A/Q4B separation
PEP	Wilcoxon	0.034	Not significant	Suggestive, underpowered
W3C	Wilcoxon	0.21	Not significant	Inconclusive

Only the RFC result survives correction for multiple comparisons. Cross-ecosystem generalization remains an open question requiring larger datasets with clean Q4A/Q4B-analogous labels.

5. Discussion

5.1 Evidence for the Q4A/Q4B Distinction

The RFC Updates vs Obsoletes split provides strong empirical support for the theoretical distinction between propagation (Q4A) and reorientation (Q4B) deployment modes:

1. **Massive effect size:** Cliff's $\delta = 0.796$ represents a near-ceiling effect, indicating almost complete distributional separation between the two relation types.
2. **Bidirectional baseline deviation:** Updates predict takeover *below* random baseline; Obsoletes predict takeover *above* random baseline. This bidirectional pattern is consistent with the Q4A/Q4B mapping and strongly disfavors simpler explanations based solely on document age or popularity.
3. **Increasing effect with signal strength:** The effect grows stronger at higher Nmin thresholds, suggesting that the Q4A/Q4B distinction becomes *cleaner* in higher-signal data rather than being an artifact of noise.
4. **Theoretical alignment:** The empirical pattern is consistent with the theoretical prediction:
 - Q4A (propagation) \rightarrow low takeover, predecessor preserved
 - Q4B (reorientation) \rightarrow high takeover, attention redirected
5. **Matched-null robustness:** The matched-null baseline ($\delta = 0.156$) is far smaller than the observed $\delta = 0.796$, indicating that document properties (age, citation volume) cannot reproduce the magnitude of the Updates/Obsolutes separation.

5.2 Connection to the Information Emergence Cycle

The Information Emergence Cycle (IEC) posits that information systems progress through four stages: Potential (Q1), Selection (Q2), Transformation (Q3), and Deploy (Q4). The Q4 stage is particularly important because it determines how transformed information re-enters the environment.

The RFC findings suggest that the Q4A/Q4B split is not merely a theoretical convenience but corresponds to genuinely different dynamics in how technical communities process revisions:

- **Q4A dynamics (Updates):** The community treats the new document as an *extension* of the old. Citations to the original remain valid and continue. The predecessor retains authority. This is propagation—new information coexists with and builds upon old.
- **Q4B dynamics (Obsoletes):** The community treats the new document as a *replacement* of the old. Citations shift to the successor. The predecessor loses authority. This is reorientation—new information redirects attention away from old.

This distinction may have broader applicability beyond technical standards:

- **Scientific literature:** Some papers extend prior work (Q4A-like) while others overturn it (Q4B-like)
- **Software versions:** Patch releases extend (Q4A) while major versions may break compatibility (Q4B)
- **Legal precedent:** Some rulings extend prior precedent (Q4A) while others overturn it (Q4B)

5.3 Mechanistic Hypotheses

Several mechanisms could explain how relation labels influence citation behavior:

1. **Direct signaling:** Authors read the “Obsoletes” label and consciously cite the successor instead of the deprecated predecessor.

2. **Indirect content effects:** Documents that obsolete predecessors may contain more substantive changes that naturally attract more citations, with the label merely reflecting this underlying property.
3. **Community coordination:** The label serves as a Schelling point for community consensus about which document is authoritative.
4. **Search and discovery:** Tools and indexes may surface “Obsoletes” successors more prominently, mechanically redirecting citations.

Our permutation baseline analysis suggests that mechanism (2) alone is insufficient—the labels carry information *beyond* what document properties alone would predict. However, we cannot definitively distinguish among the remaining mechanisms.

5.4 Limitations

1. **No preregistration:** The Updates/Obsoletes split was identified through exploratory analysis. While inclusion criteria were pre-specified independently of outcomes and falsification criteria were defined prospectively, the Q4A/Q4B hypothesis emerged from examining the data. Formal preregistration (e.g., OSF) was not conducted.
2. **Correlation not causation:** We establish strong predictive association but cannot prove that labels *cause* citation behavior. Randomized experiments are infeasible in this setting. The mechanism by which relation type predicts citation patterns—whether through author behavior, IETF deprecation announcements, community norms, or document discoverability—remains unidentified.
3. **Single ecosystem for main result:** The robust effect is demonstrated in RFCs only. Cross-ecosystem replication (PEPs, W3C) is inconclusive due to small samples and metadata limitations. Whether the Q4A/Q4B distinction generalizes beyond IETF standards is an open question.
4. **Residual confounding:** The matched-null baseline (Section 3.8) yields $\delta = 0.156$, indicating that document properties (age, citation volume) can induce modest separation but cannot reproduce the full $\delta = 0.796$ effect. We cannot rule out unmeasured confounds (document quality, author prestige, topic centrality, institutional norms) that may correlate with both relation type and takeover.
5. **RFC-specific processes:** The IETF has unique processes (formal deprecation announcements, Working Group consensus, errata systems) that may amplify the Updates/Obsoletes behavioral distinction. These RFC-specific mechanisms may not exist in other standards ecosystems.
6. **Temporal snapshot:** We analyze current citation state, not longitudinal dynamics. Takeover patterns may evolve over time, and early-stage vs late-stage citation behavior may differ.
7. **Label accuracy:** We assume Updates/Obsoletes labels are correctly applied by RFC authors. Mislabeling would attenuate effects, making our estimates conservative.

5.5 Implications

For standards bodies: The distinction between “updates” and “obsoletes” carries real behavioral consequences. Clear, accurate labeling helps communities coordinate on authoritative documents.

For knowledge evolution research: Formal metadata in technical corpora provides objective, pre-labeled data for studying how knowledge transitions propagate. The Q4A/Q4B lens offers a theoretical framework for categorizing and predicting such transitions.

For TFP framework: These results provide strong evidence consistent with the prediction that the Q4 split (propagation vs reorientation) corresponds to measurable behavioral differences in at least one real-world information system. The bidirectional baseline effects are particularly suggestive, as the direction aligns with the theoretical prediction. Generalization beyond RFCs remains an open question.

6. Conclusion

We find strong evidence that RFC “Obsoletes” relationships show systematically higher citation takeover than “Updates” relationships (Cliff’s $\delta = 0.796$, $p < 0.0001$). This effect is robust across sensitivity analyses and exhibits bidirectional deviation from permutation baselines: Updates *predict* takeover below baseline (Q4A/propagation), while Obsoletes *predict* takeover above baseline (Q4B/reorientation).

These findings support the theoretical distinction between propagation and reorientation deployment modes posited by the Information Emergence Cycle framework. The RFC corpus provides a rare natural laboratory where these theoretical categories map cleanly onto pre-existing, externally assigned labels—and the empirical pattern is consistent with the theoretical prediction.

Cross-ecosystem replication in PEPs and W3C specifications shows suggestive but inconclusive patterns, limited by sample size and metadata availability. Future work should investigate causal mechanisms, test generalization in richer external corpora, and explore longitudinal dynamics of citation takeover.

Scope and limitations: This study establishes a strong predictive association in one ecosystem (IETF RFCs) but does not prove causation—we cannot experimentally manipulate relation labels. The hypothesis was not formally preregistered, though falsification criteria were defined prospectively. The matched-null baseline ($\delta = 0.156$) shows that decade and citation volume alone cannot reproduce the observed separation ($\delta = 0.796$), but unmeasured confounds remain possible. Cross-ecosystem generalization is not established. These limitations are discussed fully in Section 5.4.

7. Reproducibility and Internal Validation

7.1 Reproducibility Infrastructure

All analyses were conducted with fixed random seeds (42) to ensure reproducibility. Code, data, and reproducibility scripts are archived on Zenodo (DOI: 10.5281/zenodo.18028423) and mirrored in the repository:

The Zenodo archive should be treated as the canonical frozen artifact for reproducing the exact version analyzed in this paper.

- `experiments/rfc_q4_split/`: Main RFC analysis
- `experiments/rfc_q4_split/reproduce_all.py`: One-command reproducibility script with sanity checks
- `experiments/paper_hardening/`: Sensitivity and baseline analyses
- `experiments/pep_cpython_replication/`: Cross-ecosystem attempts

Reproducibility checklist:

- ✓ Fixed random seeds for all stochastic analyses
- ✓ Inclusion criteria pre-specified and independent of outcome
- ✓ Sensitivity analysis across all reasonable parameter choices

- ✓ Multiple comparisons explicitly addressed (Bonferroni)
- ✓ Limitations section covers all known threats to validity
- ✓ Negative results (PEP-CPython infeasibility) documented transparently
- ✓ Effect sizes reported with confidence intervals
- ✓ Raw data and intermediate outputs preserved
- ✓ SHA-256 hashes generated for all output files

7.2 Internal Validation: Holdout Evaluation

To demonstrate that results are not artifacts of overfitting to the full dataset, we performed a stratified holdout evaluation with the following protocol:

1. **Pipeline freezing:** All analysis parameters ($N_{min} = 30$, primary test, effect size metric) were fixed *before* examining holdout data
2. **Stratified split:** Edges were randomly split 75% dev / 25% holdout, stratified by relation type to preserve class balance (seed = 42)
3. **Single evaluation:** Holdout analysis was run exactly once with no parameter tuning

Table 10: Holdout Validation Results

Metric	Dev Set (75%)	Holdout Set (25%)
n_updates	644	215
n_obsoletes	384	128
Median Updates	0.0873	0.0702
Median Obsoletes	0.7806	0.7654
Cliff's δ	0.789	0.817
p-value	6.22e-100	5.03e-37

Acceptance criteria (pre-specified): - $\delta \geq 0.5$: ✓ ($0.817 > 0.5$) - $p \leq 0.001$: ✓ ($5.03e-37 < 0.001$) - Obsoletes median > Updates median: ✓ ($0.765 > 0.070$)

Result: The effect replicates on held-out data under a frozen pipeline. Remarkably, the holdout effect size ($\delta = 0.817$) *exceeds* the dev set effect size ($\delta = 0.789$). This is atypical—most holdout validations show some degradation. The pattern suggests that the main analysis, if anything, is conservative, and argues strongly against overfitting as an explanation for the observed effect.

Note: This is an internal validation split, not a true prospective holdout—the edges were collected as one dataset. However, the frozen-pipeline protocol ensures no information leakage from holdout to analysis parameters.

7.3 Anti-Tautology Test: Normative vs Informative References

A potential concern is that the Updates/Obsolutes effect might be tautological—perhaps “Obsoletes” relationships simply involve normatively required references that must shift. To test this, we attempted to classify citations as normative (required for implementation) or informative (background only) using RFC reference section headings.

Coverage assessment: - RFCs with Normative/Informative sections: 1,329 of 9,691 (13.7%) - Citation instances classifiable: 193 of 38,787 (0.5%)

Decision: Analysis not conducted due to insufficient coverage (<50% threshold). Many RFCs—particularly older ones involved in our analyzed edges—predate the structured Normative/Informative reference format introduced in the mid-2000s.

Implication: We cannot rule out that the effect differs between normative and informative citations. However, the permutation baseline analysis (Section 3.2) partially addresses this concern: the bidirectional deviation from random pairings suggests the effect is not simply a mechanical property of reference structure.

8. Acknowledgments

The conceptual inspiration for TFP’s four-stage “Information Emergence Cycle” and the Q4A/Q4B distinction came from work on the Objective Personality System by Dave Powers and Shan Renee. TFP represents an independent engineering translation of abstract structural ideas into testable empirical predictions.

We thank the IETF RFC Editor for maintaining comprehensive public metadata that enabled this analysis, and the Python and W3C communities for transparent documentation practices that enabled replication attempts.

Use of AI tools. The author used AI assistants (Claude) as tools for code development, statistical analysis, figure generation, and assistance with drafting prose. All study design, experiments, data analysis, and scientific claims were specified, checked, and approved by the author, who takes full responsibility for the content of this manuscript.

9. Conflicts of Interest

The author declares no conflicts of interest.

10. Funding

This research received no external funding.

11. Data and Code Availability

All code, derived datasets, and figure-generation scripts supporting this paper are publicly archived on Zenodo (DOI: 10.5281/zenodo.18028423). The archive includes the one-command reproducibility entrypoint and associated sanity checks described in Section 7.1.

Repository mirror: <https://github.com/aollar/TFP-core>

License: MIT License (code), CC-BY 4.0 (data and manuscript)

How to cite: > Ollar, A. (2025). Updates Predict Low, Obsoletes Predict High Citation Takeover in Internet Standards: Reproducibility Package (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.18028423>

Data sources: - RFC metadata: IETF RFC Index (<https://www.rfc-editor.org/rfc-index.xml>) - RFC citation data: Extracted from RFC cross-reference metadata - PEP metadata: Python Enhancement Proposals repository - W3C metadata: W3C Technical Reports index

References

1. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
 2. Bradner, S. (1996). The Internet Standards Process—Revision 3. RFC 2026. Internet Engineering Task Force. <https://www.rfc-editor.org/rfc/rfc2026>
 3. Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3), 494–509. <https://doi.org/10.1037/0033-2909.114.3.494>
 4. Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.
 5. IETF Datatracker. (2024). Document Relationship Types. <https://datatracker.ietf.org/doc/>
 6. Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1), 50–60.
 7. Python Software Foundation. (2024). PEP 0—Index of Python Enhancement Proposals. <https://peps.python.org/>
 8. Python/Cython Repository. (2024). GitHub. <https://github.com/python/cpython>
 9. RFC Editor. (2024). RFC Index. <https://www.rfc-editor.org/rfc-index.html>
 10. Romano, J., Kromrey, J. D., Coraggio, J., & Skowronek, J. (2006). Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys? *Annual Meeting of the Florida Association of Institutional Research*.
 11. W3C. (2024). Web Platform Specs. <https://www.w3.org/TR/>
 12. Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83. <https://doi.org/10.2307/3001968>
 13. Ollar, A. (2025). Updates Predict Low, Obsoletes Predict High Citation Takeover in Internet Standards: Reproducibility Package (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.18028423>
-

Appendix A: Information Emergence Cycle Background

The Information Emergence Cycle (IEC) is a four-stage model describing how information systems process and deploy new information. The framework posits that information flows through stages of accumulation, selection, transformation, and deployment (see Acknowledgments for conceptual origins).

Table A1: Information Emergence Cycle Stages

Stage	Name	Function	Analogy
Q1	Potential	Accumulate latent possibilities	Gathering raw materials
Q2	Selection	Choose specific configurations	Deciding what to build
Q3	Transformation	Process selected configurations	Building/manufacturing

Stage	Name	Function	Analogy
Q4	Deploy	Release transformed information	Shipping the product

The Q4 (Deploy) stage is unique in that it interfaces with the external environment. This stage admits two distinct modes:

- **Q4A (Propagation):** Deployment extends existing structure. The new information is *additive*—it coexists with predecessors, builds upon them, and does not invalidate them. In citation terms, both old and new documents remain valid references.
- **Q4B (Reorientation):** Deployment replaces existing structure. The new information is *substitutive*—it supersedes predecessors and redirects attention away from them. In citation terms, the new document becomes the authoritative reference.

This paper provides strong evidence consistent with the prediction that the Q4A/Q4B distinction corresponds to measurable behavioral differences in the RFC ecosystem. Whether this generalizes to other technical standards ecosystems or information systems remains an open question.

Appendix B: Supplementary Tables

Table B1: Full Sample Characteristics

Characteristic	Updates (Q4A)	Obsoletes (Q4B)
Total edges	1,266	921
Edges with ≥ 30 citers	859	512
Median total citers	58	71
Median old-doc citers	42	31
Median new-doc citers	12	48
Median age difference (years)	4.2	6.1

Table B2: Bootstrap Confidence Intervals

Statistic	Updates (Q4A)	Obsoletes (Q4B)
Median takeover	0.083 [0.07, 0.10]	0.773 [0.75, 0.80]
Mean takeover	0.152 [0.14, 0.17]	0.698 [0.67, 0.72]
Cliff's δ	-	0.796 [0.77, 0.82]

Note: 95% bootstrap CIs with 10,000 resamples.