

Supplementary Materials Guide

Product Tags and the Observability of Marketplace Enforcement: Evidence from DSA Transparency Data

Bahadır Yüzbaşı and Serkan Benk

Version 1, 27 October 2026

1 Scope and interpretation

The nine CSV tables accompany the article and preserve the completed analysis outputs without changing their values or column names. They describe the five-marketplace snapshot for September–November 2025. The 20,528 aggregation cells have a source-frequency total of 346,993,630 statements. This total is not a count of independently observed people, sellers, products or incidents.

All-reported-types and product-tagged records define distinct content populations. Domain indicators may overlap. The multiple-domain outcome means at least two positive restriction domains and is not a severity scale. Automation comparisons are descriptive, pair-specific quantities; they are not hypothesis tests or causal estimates.

2 Inventory of supplementary tables

Label	Title	Main contents
S1	Platform populations and domain rates	Platform–population totals, aggregation-cell counts, positive-domain frequencies and percentages.
S2	Platform–month domain rates	The corresponding totals and rates by calendar month.
S3	Joint restriction bundles	All 16 possible combinations of the four restriction domains, including observed zero counts.
S4	Decision-automation profiles	Frequencies and domain rates for fully, partially and non-automated groups.
S5	Detection–decision profiles	Frequency by the joint automated-detection and decision-automation labels.
S6	Product-tag capture	Product-tagged positive frequency, full-population positive frequency and their ratio by domain.
S7	Common-support flows	Pair-specific eligible and retained frequencies, arm counts, shared-stratum counts and support coverage.
S8	Standardized rates and differences	Both arm-specific rates, frequency counts, support diagnostics, standardized rates, differences and undefined-status indicators.
S9	Daily domain counts and rates	The complete 91-day platform calendar, both content populations and zero-volume status.

Each file is named `Table_S[number]_[descriptive name].csv`. The accompanying checksum manifest records the exact filename and SHA-256 hash for each file. Renaming the original tables for submission does not change their contents.

3 Units, group codes and missing values

Frequency and event-count columns contain source statement frequencies. Cell counts refer to aggregation cells. Neither is a survey weight. Columns ending in `_pct` use percentage units; `difference_pp` uses percentage points.

`all_reported_types` denotes the full recorded content population; `product_type_true` denotes records whose product-content flag is true. FULLY, PARTIALLY and NOT_AUTOMATED correspond to F, P and N in the article. Pair order determines the subtraction: F–P means the F rate minus the P rate.

An empty numeric cell is undefined, not zero. In Table S6, UNDEFINED_ZERO_DENOMINATOR indicates no positive full-population denominator. In Table S8, NOT_IDENTIFIED_NO_COMMON_STRATA indicates that the two arms do not share an eligible reporting stratum. A reported numeric zero remains an observed zero for the specified quantity.

Tables S7 and S8 describe overlapping pair-specific populations. Their frequencies must not be summed across contrasts. Support requires positive frequency in each arm within the same platform, category, source creation day, source type and automated-detection status. There is no additional minimum-count or outcome-based selection rule. Both arms and all five outcomes use the same frozen joint-frequency weights for each comparison.

Amazon’s published source volume is zero on 20–21 October 2025. These dates remain in Table S9, with undefined statement-based rates. They were not imputed or excluded from the calendar.

4 Source-field mapping

The following mapping links the article’s readable variable names to the source schema. It documents the existing definitions and adds no transformation or recoding.

Article terminology	Exact source definition
Reporting day	created_at, interpreted at day resolution; not the archive filename or application date.
Statement frequency	count, retained as an integer source frequency.
Product-tagged population	CONTENT_TYPE_PRODUCT equals true.
Visibility restriction	At least one field beginning DECISION_VISIBILITY_ equals true.
Account restriction	decision_account is nonempty.
Monetary restriction	decision_monetary is nonempty.
Service restriction	decision_provision is nonempty.
Multiple domains	At least two of the preceding four domain indicators are positive.
Decision automation	automated_decision; source values have prefix AUTOMATED_DECISION_.
Exact matching fields	platform_name, category, created_at, source_type, automated_detection.

Blank restriction fields contain no positive record; they do not establish the absence of a real restriction. Source normalization of Boolean fields limits distinctions between an explicit false value and a value normalized to false. Source-cell identifiers identify aggregation cells, not individual statements or people.

5 Data source and reproduction

The source is the European Commission’s DSA Transparency Database: <https://transparency.dsa.ec.europa.eu/explore-data/download>. The study uses the archived aggregated-simple exports for September–November 2025.

A separate replication archive contains the frozen analytical data, exact support memberships and weights, the original nine result tables as references, and a portable run.py script. It requires Python 3.9 or later and uses only the standard library. From the extracted directory, run:

```
python3 run.py --check
python3 run.py
```

The first command verifies input hashes and snapshot totals. The second reconstructs the nine tables in results/tables and checks them against the supplied reference files. It refuses to overwrite an existing results run. No random seed, package installation, model fitting, resampling or hypothesis testing is involved.

The replication archive contains the selected analytical snapshot. The much larger complete monthly source exports and targeted archive-audit files remain available through their source/provenance records rather than being duplicated in the small replication archive.