



METHODOLOGY DOCUMENTATION

Buyer & Broker Matching Methodology

Matching registry retirement records to known buyers and brokers

DATE	TYPE	PREPARED BY
September 2026	SQL + Python	AlliedOffsets Product Team

Overview

When a carbon credit is retired, the registry provides only free-text fields describing who retired it and why. This methodology converts that raw text into a clean, trusted buyer (the organisation that ultimately claimed the offset) and, where applicable, a broker (the intermediary that carried out the retirement on the buyer's behalf). Matching happens in two stages: an automated daily process that stamps known buyer and broker names onto new transactions using a set of trusted rules, and a manual process, run three times a week, that proposes matches for transactions the automated stage could not resolve. Proposed matches are held in a staging area until a human confirms them, at which point they are added to the trusted reference list for future use.

Data Inputs

Field / Parameter	Description	Source / Notes
Account Holder	The raw registry-provided text identifying who holds or retired the credits.	Registry retirement record
Retirement Details	Free-text notes accompanying a retirement, often naming the buyer, broker, order reference, or reason for retirement.	Registry retirement record
Project, Retirement Date, Credits Retired	The specific project, the date credits were retired, and the volume of credits retired in that transaction. Used to recognise repeat transactions that share the same fingerprint as a previously matched one.	Registry retirement record
Reference Mapping List	The trusted, human-maintained list of raw text patterns and their corresponding clean buyer and broker names. This is the single source of truth all automated matching is checked against.	Human-maintained reference sheet, loaded into the database
Known Buyer Names	The distinct list of buyer names already confirmed in the reference mapping list, used as a lookup for word-boundary text matching against new, unmatched transactions.	Derived from the Reference Mapping List
Known Broker Names	The distinct list of broker names already seen in the transaction data, excluding a small set of names that are deliberately excluded because they produce noisy matches.	Derived from historical transaction data
Alias and Merger Overrides	A fixed set of name variants and historical company mergers (for example, alternate spellings of the same broker, or two brokers that have since merged into one) that are mapped to a single clean broker name.	Maintained directly within the manual matching process
Cutoff Date	The timestamp of the last time the manual matching process was run. Only transactions retired, created, or changed since this date are considered for manual matching.	Stored value, updated after each manual run



Field / Parameter	Description	Source / Notes
Registry Scope	The set of registries the automated matching rules are applied to for a given run.	User-supplied parameter

Logical Flow

1	Maintain the trusted reference list A human-maintained reference sheet records every raw text pattern seen so far and the clean buyer and/or broker name it corresponds to. This sheet is the source of truth for all automated matching and is loaded into the underlying reference table before each run.
2	Match on exact account holder and retirement text New transactions are compared against the reference list for an exact match on both the account holder text and the retirement details text. Where both fields match a known entry exactly, the corresponding buyer and broker are applied.
3	Match on project, date, and volume fingerprint Transactions are also compared against previously matched transactions that share the same project, the same retirement date, and the same total volume of credits retired. A match on all three is treated as strong evidence it is the same underlying transaction, and the known buyer and broker are carried across.
4	Apply fixed platform rules A small number of tokenised carbon-credit platforms are recognised directly from keywords in the retirement text (for example, retirements routed through a specific token wrapper), and are assigned a fixed buyer name without needing a reference list entry.
5	Match on text patterns Where the retirement text follows a recognisable pattern rather than an exact string, a set of approved text patterns is checked against the retirement details field. A pattern match applies the associated buyer and broker.
6	Combine and apply automated matches Results from all active automated rules are combined per transaction, duplicates are removed, and the resulting buyer and broker values are written back onto the transaction record. This automated pass runs on every relevant transaction and only ever produces a match where the text has been seen before.
7	Identify transactions still unresolved On a fixed schedule (Monday, Wednesday, Friday), the process identifies every transaction retired, newly created, or changed since the last run that still has no buyer or broker assigned, alongside those that were already resolved by the automated pass.
8	Fuzzy-match against known names Each unresolved transaction's account holder and retirement text is scanned for a whole-word match against the list of known buyer names and, separately, known broker names. A match assigns the corresponding name as a proposed buyer or broker.
9	Apply alias and merger overrides A fixed set of alternate spellings, common abbreviations, and historical company mergers is applied on top of the fuzzy match, so that different text variants of the same underlying broker are all resolved to one current, clean name. One broker name is treated as a special case and always overrides any other match, because it is known to retire credits on behalf of other brokers.



10

Extract buyer names from marketplace order text

For retirements made through a specific marketplace where the buyer is not yet in the known list, the text following the word 'for' in the retirement details is treated as a candidate buyer name. Text that looks like an individual's name rather than a company is discarded; legal suffixes such as Ltd, Inc, or GmbH are trimmed from the end of company names that are kept.

11

Discard transactions already covered by the reference list

Before anything is proposed for review, any transaction whose account holder and retirement text pair already exists in the reference list is removed, so the same text is never proposed twice.

12

Write proposed matches to staging

The remaining proposed matches, along with transactions that could not be matched at all, are written to a staging area, tagged as either matched via the known list or still unresolved.

13

Human review and reference list update

A person reviews the staging area, confirms or corrects the proposed buyer and broker names, and adds approved entries back into the trusted reference list. The cutoff date is then moved forward to the date of this run, so the next run only looks at transactions retired or changed after this point.

Key Calculations

Retired credit volume aggregation

For each unique combination of account holder, retirement text, retirement date, and project, the total number of credits retired is summed across all underlying records. This summed volume is what is compared against previously matched transactions in the project/date/volume fingerprint match, rather than comparing individual line items.

Assumptions

- Retirement text containing the word 'toucan' or the token 'C3T' is always attributed to that platform as the buyer, regardless of any other text present, and without a broker being assigned.
- A match on project, retirement date, and total credit volume is treated as sufficient evidence that a new transaction is the same underlying retirement as a previously matched one; it does not require any text similarity.
- Text pattern matching assumes the pattern appears as a substring anywhere in the retirement details field, which means two unrelated names sharing a common substring could in principle produce a false match.
- The buyer name 'Carbon Neutral' is deliberately excluded from the automated buyer name list used for fuzzy matching, to avoid over-broad matches on a very generic term.
- The broker names 'ACT', 'Clear', 'Wren', and 'Carbon Neutral' are excluded from the distinct broker list pulled from historical data, again to avoid noisy or over-broad matches; some of these are reintroduced later via specific, narrower rules.
- A fixed set of alias and merger overrides is applied after the general fuzzy match, covering known alternate spellings and company mergers (for example, two historical broker names that have since combined into one current entity, or a common abbreviation that maps to a longer registered name). These overrides are maintained directly in the matching code rather than in a data table, so adding a new alias currently requires a code change rather than a data update.
- One broker name is explicitly treated as an override that takes priority over any other broker match on the same transaction, because that broker is known to carry out retirements on behalf of other brokers.



- One broker name match is deliberately written to exclude a longer, unrelated name that contains it as a substring, to avoid misattributing that unrelated entity's retirements.
- When extracting a buyer name from marketplace order text, short two-word capitalised text with no company-type keyword (such as 'Ltd', 'Foundation', or 'Group') is assumed to be an individual's name and is discarded rather than proposed as a buyer.
- Legal suffixes (for example Ltd, Inc, GmbH, LLC) are only stripped when they appear at the very end of an extracted buyer name, so that suffixes appearing mid-name are left untouched.
- The manual matching process only considers transactions retired, newly created, or changed since the last recorded run date, on the assumption that everything before that date has already been reviewed.

Limitations

- All automated and fuzzy matching rules can only recognise buyers and brokers that already appear, in some form, in the reference list or historical transaction data. A genuinely new buyer or broker that has never been seen before will not be identified automatically and requires a human to add it.
- The manual matching process runs on a fixed schedule (Monday, Wednesday, Friday) rather than continuously, so a newly retired credit may not be matched, even provisionally, for up to a few days.
- The human review step is a required gate before any proposed match is trusted; the speed of this review, rather than the matching logic itself, is typically the main constraint on how quickly new buyers and brokers become fully recognised.
- Buyer name extraction for marketplace order text is a best-effort heuristic based on capitalisation and keyword presence; it will occasionally misclassify an individual as a company or vice versa, and any result from this step is treated as unconfirmed until reviewed.

Output

Each transaction ends up with a main buyer name and, where applicable, a broker name attached to it. Transactions resolved by the automated pass are considered trusted and applied directly. Transactions resolved only by the manual, fuzzy-matching pass are held in a staging area as proposed matches until a human confirms them, at which point they are folded into the trusted reference list for all future runs.