

METHODOLOGY DOCUMENTATION

Methodology Standardization

Normalizing registry methodology codes and names into a standardized methodology reference

DATE September 2026	TYPE Python	PREPARED BY AlliedOffsets Product Team
-------------------------------	-----------------------	--

Overview

Carbon credit registries reference the methodology used to quantify a project's emissions reductions using their own codes and naming conventions (for example 'VM0007', 'ACM0002', or 'AMS-I.D.'). and projects frequently cite more than one methodology at once, separated by commas or other delimiters. AlliedOffsets standardizes these raw methodology references into a consistent internal naming and versioning scheme, so that clients can compare, filter, and analyze projects by methodology consistently regardless of how the originating registry labelled it.

Data Inputs

Field / Parameter	Description	Source / Notes
Raw methodology text	The methodology code or name exactly as published by the registry for a project, which may contain one or several methodologies combined into a single text field.	Registry data feeds, captured during data extraction for each registry.
Registry	The name of the originating registry, recorded alongside the raw methodology so unmapped entries can be reviewed in context.	Internal registry identifier associated with each data source.
Methodology mapping table	The internal reference table holding every raw methodology value encountered so far, its standardized name, and its version, since a single raw name can resolve to multiple standardized methodology/version entries.	AlliedOffsets internal database, seeded with a large reference list of known methodology codes (e.g. ACM0002, VM0007, AMS-I.D.).

Logical Flow

1

Load the current mapping table

Every existing raw methodology entry is loaded from the internal mapping table into memory. Because one raw methodology name can correspond to multiple standardized methodology/version pairs, each raw name is loaded as a list of possible standardized entries rather than a single value.

2

Check for a direct whole-text match

The full raw methodology text for a project is first checked as-is against the mapping table, in case it already exactly matches a known single methodology entry.

3

Split combined methodology references

If no direct match is found, the raw text is checked for common separators (comma, plus sign, semicolon, or line break) that indicate multiple methodologies were combined into one field. When a separator is found, the text is split into individual methodology references.

4 Map each individual methodology

Each split-out methodology reference is trimmed of extra whitespace and looked up individually in the mapping table, with the original combined text retained for context if the entry needs to be reviewed.

5 Flag unmapped methodologies

Any methodology reference — whole or split — that is not already known is automatically added to the mapping table with no standardized value assigned yet, tagged with the registry it came from and the original combined text it was extracted from, and a notification is sent to the team responsible for classification.

6 Validate before publishing

Before a batch of projects is finalized, every methodology name attached to a project is checked against the full set of recognized standardized methodology names. Any project referencing a methodology that has not yet been resolved to a standardized entry is flagged as a validation error.

Key Calculations

Multi-methodology splitting

When a project's raw methodology field contains more than one methodology combined together, the text is divided using a fixed set of recognized separators (comma, plus sign, semicolon, line break) in that priority order. Each resulting piece is mapped independently, and the project is associated with the full list of standardized methodology/version entries that result — not just a single value.

Assumptions

- The methodology reference list is seeded with a large catalogue of known raw methodology codes (for example ACM0002, VM0007, AMS-I.D.) along with descriptive and scoring information, giving new registry data a head start toward a match before anything needs manual classification.
- Only four separators — comma, plus sign, semicolon, and line break — are recognized as indicating multiple combined methodologies. A registry using any other delimiter would have its combined text treated as a single (likely unmapped) methodology reference.
- If splitting a combined methodology reference produces no successfully mapped pieces, the system falls back to treating the entire original combined text as a single methodology lookup, in case the whole string exactly matches a previously catalogued combined entry.
- A standardized methodology entry with both its name and version unset is treated as equivalent to no valid mapping, even if a database row technically exists for it.

Limitations

- Newly encountered methodology references — whether whole or split out from a combined field — are not automatically classified and require manual review before resolving to a standardized name and version.

- The splitting logic assumes registries use one of four known separators for combined methodology fields; unusual or inconsistent formatting from a registry could result in incorrect splitting or a failed match.
- The process standardizes the methodology's name and version; it does not independently verify that a methodology code is still valid, current, or applicable to the project type it is associated with.
- Because a single project can cite multiple methodologies, the standardized output is a list rather than one value — downstream consumers of this data need to account for projects having more than one associated methodology.

Output

Each project record is enriched with a list of standardized methodology entries (name and version) resolved from its raw registry methodology text, which can be used consistently for filtering, comparison, and reporting across all registries on the AlliedOffsets platform.