Lead Attribution Engine

Capturing the true source of every lead for a US roofing and solar contractor

Hugo · October 2026 · Case study

Summary

A major US roofing and solar contractor was spending heavily on paid ads, but its CRM could not say which channel produced which lead. Paid Google leads went Unattributed, phone calls arrived with no source, and the main estimate form was not connected to the CRM at all.

I designed and built a Lead Attribution Engine that records where every lead came from at the moment of first contact, by web form or phone call, and writes it into the CRM in a form the attribution workflows can credit. Web-form capture went live and was verified site-wide in July 2026. Phone tracking is in a single-market pilot ahead of full rollout.

At a glance

Client	Major US roofing and solar contractor
My role	Project owner, lead engineer and integrator
Status	Web forms: live site-wide (July 2026) · Phone: pilot in one market
Stack	JavaScript, HubSpot Forms V4 API, HubSpot workflows and properties, Webflow custom code
Integrations	HubSpot CRM, Leap (formerly JobProgress), CallTrackingMetrics, Google Ads, Microsoft Ads, Meta Ads, MNTN, FlowCode, Roofr

Key results

- **30 lead forms → 2.** Each now carries 6 hidden tracking fields, and a form change is made once instead of 30+ times.
- **Site-wide source capture.** One script fills the tracking fields on every form, with no per-page setup.
- **Lost leads recovered.** The main Free Estimate form, which sent nothing to the CRM, now creates contacts with full source data.
- **Phone source verified.** The call tracker's source reached the CRM field on 40 of 40 recent phone contacts.
- **Silent form failures ruled out.** Every value the script can write was checked against the CRM's 215-option picklist.

The challenge

Leadership wanted cost per lead, cost per booked appointment, customer acquisition cost (CAC) and return on ad spend (ROAS) by channel and campaign. Spend covered Google, Bing, Meta and MNTN streaming TV ads, plus organic search, Google Business Profile, direct mail and a referral program. The lead data could not support those numbers.

The CRM was crediting leads from one wrong field. The attribution workflows read only `lead_source` and map it to a channel by prefix. Paid Google leads, confirmed as paid by HubSpot's own analytics, were written as a bare "Google". That value maps to no channel, so they went Unattributed.

Most leads had no source at all. Phone calls make up most lead volume, and they arrived in the CRM as "Offline".

Organic and widget leads looked like Direct. HubSpot's tracking script loaded too slowly for some organic visitors, and the third-party Roofr quote widget does not pass HubSpot's tracking cookie. Both cases were recorded as Direct traffic.

Some leads never reached the CRM. The main Free Estimate form, the target of every Free Estimate button on the site, was a Webflow form with nothing behind it. An audit of the automation tools confirmed none of its submissions became contacts.

The site made fixes expensive. An audit found 30 distinct lead forms across 100 of the site's 813 live pages. None of the location-page forms had tracking fields, and 11 cities shared one identical copy with no city value. With 18 pages embedding the same form two to five times and no shared page templates, every form change meant editing pages by hand.

The approach

The engine is a four-stage pipeline. Every stage depends on clean, specific data at capture, so that is where the build focused.

Capture → Channel stamp → Deal attribution → Reporting

1. **Capture.** A web form or tracked call writes the real source into `lead_source`. Web leads also store the Google click ID (`gclid`) and UTM parameters.
2. **Channel stamp.** A HubSpot workflow maps `lead_source` to a channel by prefix and date-stamps the contact.
3. **Deal attribution.** When a qualifying deal is created, a second workflow sets its Attributed Channel using a **14-day last-touch model with a sticky fallback**. The deal goes to the channel that touched it in the previous 14 days, otherwise to the contact's last-known channel, and only then to Unattributed.
4. **Reporting.** Deals roll up by channel and campaign for the marketing team and the data analyst.

The solution

1. Web-form source capture

Two forms instead of thirty. One Residential form serves the location, Free Estimate, and windows, doors and siding pages. One Commercial form serves the commercial contact, coatings, membrane and metal roofing pages. Each carries 6 hidden fields: `gclid`, `utm_source`, `utm_medium`, `utm_campaign`, `utm_content` and `lead_source`.

One site-wide capture script. A single script in the site's global head code reads the landing URL, click IDs and referrer, then fills the hidden fields before submission. It sets `lead_source` by this priority:

Priority	Signal	Value written
1	Google click ID, or Google with a paid medium	`Google Ads | Other`
2	Bing with a paid medium	`Bing | Other`
3	Facebook or Meta with a paid medium	`Meta - Other`
4	Search-engine referrer, no paid signal	`Organic Search`
5	Any other external website	`Referral`
6	Direct or unknown	Left blank

Two hard rules.

- Never write a generic "Google" or "Website". A blank field is honest; a generic value is misleading.
- Always store the click ID and UTMs as an audit trail, even when `lead_source` is set.

Channel and campaign kept apart. `lead_source` carries the channel only. Campaign names live in `utm_campaign` and a dedicated campaign field.

2. Phone-call source capture

Dynamic Number Insertion (DNI). CallTrackingMetrics (CTM) shows each visitor a tracking number that matches their source, and every number routes to the central lead center. The number's tracking source is written to the contact and copied into `lead_source`.

Values the engine can credit. Each number writes a specific value from the engine's taxonomy, such as `Google My Business - <city>`, `Organic Search` or `MNTN`, never a bare "Google".

An organic-focused pilot. Paid channels were already credited through HubSpot's ad-interaction and conversion tracking. The pilot therefore covers Google Business Profile website clicks, Google organic, Bing organic, organic social, and an MNTN streaming TV and QR number.

Rollout sized. About 28 location pages across each traffic source comes to roughly 325–540 numbers, in pools of 3–5 per source so high traffic does not cause conflicts.

QR codes that never need reprinting. The QR destination URL carries a UTM tag set in FlowCode, so CTM shows the right number on scan. The tag can change later without reprinting the code.

3. Supporting work

- **Decision dashboards.** One single-source-of-truth Linear dashboard per track (forms, phone and `lead_source` decisions), each with requirements, ordered tasks, a decisions log and a change log. Repository rules and an automated weekday audit kept them current.
- **Form and site audit.** Documented duplicate forms, missing templates, dead pages, broken CSS, dead calls to action and SEO issues, with a recommendation to migrate to templates before rollout.
- **Roofr widget (phase 2).** Two options are under evaluation: pull submissions server-side through the Roofr API and add the source, or capture the source on the page that embeds the widget.

Results

Web-form capture is live and verified across the whole site. Phone tracking is in its pilot, and end-to-end verification of the deal-side Attributed Channel is still to come.

Area	Before	After
Paid Google web leads	Bare "Google", Unattributed	Click ID, UTMs and `Google Ads | Other` captured; credited to Google Search (PPC)
Organic web leads	Often recorded as Direct	Classified from the referrer as `Organic Search`
Phone leads	"Offline", no source	Specific, creditable source per tracked number (pilot)
Free Estimate form	Not connected; leads lost	HubSpot form with full source capture
Lead forms	30 forms, no tracking fields, cloned per city	2 shared forms with 6 hidden tracking fields
New campaigns	Blank until added to the list by hand	Credited automatically through a channel catch-all
Form changes	Repeated on 30+ copies	Made once, in 2 places

What this means for the business

- **No silent lead loss.** Every value the script can produce exists in the CRM's dropdown, so no submission can be rejected.
- **Fewer Unattributed deals.** Paid, organic and referral leads now arrive with a value the engine can credit. This is what makes channel-level CAC and ROAS reporting trustworthy.
- **Every attribution is checkable.** Click IDs and UTMs are stored on every web lead, so any credit can be compared with HubSpot's native source.

What's next

- Roll out phone tracking beyond the pilot market.
- Verify the deal-side Attributed Channel end to end.
- Bring Roofr widget leads into the engine (phase 2).

Engineering highlights

A closed set of output values.
Problem: In live testing, HubSpot rejected the entire submission (HTTP 400) when a dropdown received a value that was not an existing option. The Submit button simply did nothing, and the lead was lost silently.
Fix: Paid leads now write fixed channel catch-alls, and campaign detail moved to its own field. I then audited every possible output against the live picklist, which makes the failure impossible rather than unlikely.

Built for HubSpot's current forms.
Problem: Legacy form callbacks never fire on forms built in HubSpot's new editor.
Fix: I rewrote the script to listen for the new forms' ready event and write values through the Forms V4 API. The listener must register before the HubSpot embed script, because the event fires only once.

Tests that prove the script ran.
Problem: HubSpot pre-fills fields from matching URL parameters by itself, so UTMs can appear even when the script is absent.
Fix: Every smoke test checks the derived `lead_source`, which never appears in the URL and so proves the script ran.

Root-causing a hidden blocker.
Problem: The tracking properties would not appear in the form editor. Brand assignment was first blamed.
Fix: I traced it to a property flag set when the properties were created through the API, and corrected the plan.

A verified data bridge. I confirmed the call tracker's source field is copied into the field the engine reads (40 of 40 recent phone contacts), and checked every planned phone value against the live picklist before wiring anything.

Safe delivery. A 40-case automated test suite, regression guards for the exact failure shapes found in production, one repository source of truth for the script with a paste-ready browser build, and a pilot on one production page before the site-wide rollout.

My role

I own the project and led it from discovery through delivery.

- **Discovery.** Wrote the original investigation into why leads showed up as Direct. That investigation became this project.
- **Requirements and decisions.** Turned the attribution spec into tracked requirements and ran the decisions on the source field: dropdown versus text, catch-all values, referral, organic grouping and the campaign-field split. Each decision is recorded with its reasoning.
- **Build.** Wrote and tested the capture script, designed the 2-form consolidation, deployed the script site-wide, and am rolling the forms out across the site.
- **Phone tracking.** Designed the DNI pilot, sized the full rollout, set the CTM tracking values and defined the QR setup.

- **Audits.** Audited the forms and site, the automation tools, the CRM picklist and the call-tracking data bridge.
- **Stakeholders.** Coordinated with the marketing lead and the data analyst through the Linear dashboards, so each track always showed its next step, owner and blockers.

Appendix A: Lead sources

Source	Type	How it is captured	Value in `lead_source`	Credited channel
Google Ads (Search)	Paid	Click ID or `utm_source=google` with a paid medium; HubSpot ad-interaction tracking; tracked call numbers	`Google Ads | Other` (web); `Google Ads Calls` (phone)	Google Search (PPC)
Google Ads (Solar)	Paid	Same as Google Ads	`Solar Google Ads …`	Google Search (PPC)
Google Local Services Ads	Paid	Tracked call numbers	`Google LSA - <city>` / `Google LSA | Other`	Google LSA
Google Business Profile	Organic	Static tracked number for profile website clicks, plus direct listing calls	`Google My Business - <city>`	Google Business Profile
Microsoft (Bing) Ads	Paid	`utm_source=bing` with a paid medium	`Bing | Other`	Bing
Meta (Facebook / Instagram)	Paid	`utm_source=facebook` or `meta` with a paid medium; HubSpot ad-interaction tracking	`Meta - Other`	Meta
Organic search (Google, Bing)	Organic	Search referrer with no paid signal (web); dedicated DNI numbers (phone)	`Organic Search`	Organic Search
Organic social	Organic	Dedicated DNI number	`Organic Social`	Not credited yet (accepted for the pilot)
MNTN streaming TV / QR	Paid	Dedicated tracking number; QR URL tagged in FlowCode	`MNTN`	MNTN
Direct mail	Paid	Mailer-specific values	`Mailers - Radius` / `Mailers - Target`	Direct Mail
Broadcast	Paid	Asset-specific tracked numbers	`Broadcast - <asset>`	Broadcast
Website referral	Organic	External, non-search referrer	`Referral`	Referral

Source	Type	How it is captured	Value in `lead_source`	Credited channel
Customer referral program	Referral	Referral platform writes the value on every submission	`Referral Program`	Referral program
Roofr quote widget	Mixed	Phase 2: Roofr API or on-page capture	Same rules as web forms	Per the source
Direct / unknown	None	No click ID, UTM or external referrer	Blank, never "Direct" or "Website"	Last-known channel via sticky fallback, otherwise Unattributed

Appendix B: Systems

System	Category	Role in the engine
HubSpot	CRM and marketing automation	System of record for contacts and deals. Hosts the two forms, the `lead_source` picklist and tracking properties, the Channel Stamp and Deal Attribution workflows, and reporting lists
Leap (formerly JobProgress)	Job-management CRM	System of record for roofing and other non-solar jobs. Roofing revenue (total job price at signing) feeds CAC and ROAS reporting
HubSpot Forms V4 API	API	Writes the hidden fields through the form's ready event before submission
HubSpot Properties API	API	Used to audit the live `lead_source` picklist (215 options) against every value the script can emit
CallTrackingMetrics	Call tracking / DNI	Swaps numbers by source, routes calls to the lead center, writes the call's source to the contact
Webflow	Website CMS	Hosts the site; global head code delivers the capture script to every page
Google Ads, Microsoft Ads, Meta Ads	Ad platforms	Supply the click IDs and UTM tags the script reads, and spend data for reporting
MNTN	Streaming TV ads	Paid channel with its own tracking number; used as the QR test
FlowCode	QR codes	Sets tagged QR destination URLs so scans show the right tracking number
Roofr	Instant-quote widget	Lead source joining the engine in phase 2
Customer referral platform	Internal app (built by me)	Writes the referral-program value on every submission
Zapier	Automation	Audited; confirmed no form data reached the CRM through it
Marketing performance dashboard	Reporting	Downstream consumer: cost per lead, cost per booked appointment, CAC and cohort ROAS by channel, then campaign

Appendix C: Glossary

- **Attribution:** crediting a lead or deal to the marketing channel that produced it.
- **CAC (customer acquisition cost):** marketing spend divided by customers won.
- **DNI (Dynamic Number Insertion):** showing each visitor a phone number tied to their traffic source.
- **gclid:** Google's click ID, added to the URL when someone clicks a Google ad.
- **Last-touch model:** credits the deal to the most recent channel that touched the lead.
- **ROAS (return on ad spend):** revenue divided by ad spend.
- **UTM parameters:** tags added to a URL to identify the source, medium and campaign of a visit.