



SCRAPPOINT · WHITE PAPER

The Hidden Cost of Outdated B2B Data

What a large database is really costing your agency — the decay math, the hidden operational cost, and the process that fixes it.

For founders & agency leaders

Published October 2026 · Scrapoint Research
scrapoint.com

01

Executive Summary

2.1% / mo

B2B contact-data decay rate [2]

22-30% / yr

records sliding to inaccurate [2]

\$12.9M / yravg cost of poor data quality,
Gartner [1]

The problem. Agencies measure their prospect database by size. The number on the file, however, says nothing about how much of it is still true. B2B contact data decays at roughly 2.1% per month — about 22–30% a year at baseline, and faster when volatile fields like job title are involved [2]. A database left untouched is losing tens of thousands of usable records a year, silently.

Why it matters. For an agency, the data is the product. Decay does not show up as a line item; it shows up as researcher hours, QA, rework, campaign waste, client revisions, and damaged sender reputation. Gartner estimates poor data quality costs the average organization \$12.9 million a year [1].

Who is affected. Founders and MDs of lead-generation, advertising, and marketing agencies — because the cost lands on margin and on client retention, not on an obvious data budget.

Key findings.

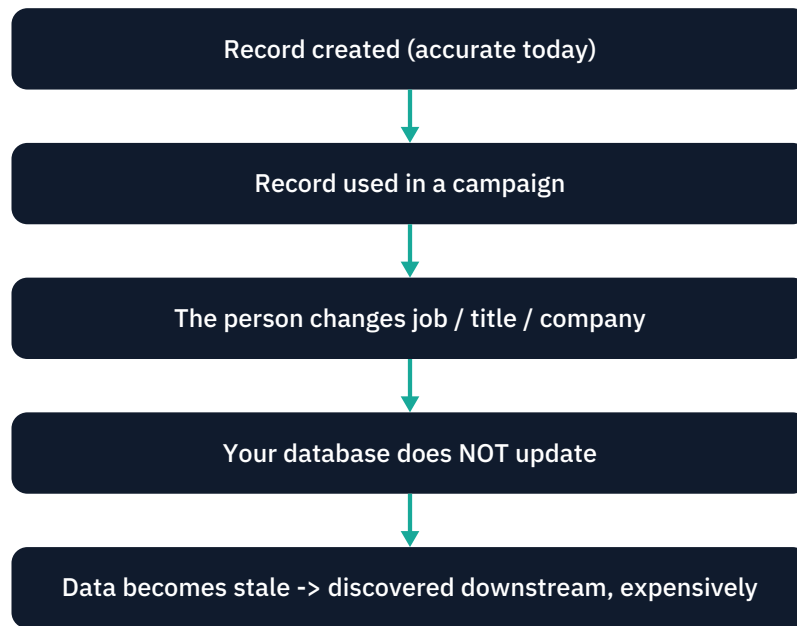
1. Volume != value: a 100,000-record database can hold far fewer usable records.
2. The fastest-decaying field (job title) is the one agencies target on [2].
3. Adding researchers scales the problem, not just the capacity.
4. One-time cleanups cannot fix a continuously-decaying asset.

Recommended approach. Treat the database as a system — collect, normalize, store, and refresh on a priority cadence — not as a file you finish.

02

The Data Decay Problem

Data decay is a cycle, not an event:



The cost is not the decay itself — it is discovering the decay after you have spent money acting on the record. Median US tenure is now 3.9 years, the shortest since 2002, and about a quarter of workers change jobs in a given year [3]. Your file does not move when they do.

03

What Changes in a B2B Record?

Data field	Possible change	Relative decay speed
Job title	Promotion, role change	Fastest [2]
Company	Job change, acquisition	Fast
Email	Job move, domain change	~3.6% / month [2]
Location	Relocation	Slow
Company size	Growth / contraction	Moderate
Department	Reorg, role change	Moderate
ICP fit	Follows title & company	Moves with the above

The takeaway: records do not fail uniformly. The fields that define targeting — title, seniority, company — are among the first to go.

04

The Real Cost

The true cost of a database is operational, not acquisition:

Actual data cost = research + QA + rework + management overhead

- **Research** — collecting and re-verifying records.
- **QA** — a second team re-checking the first.
- **Rework** — fixing what shipped wrong.
- **Management** — coordinating people, formats, and versions.
- **Client revisions** — the errors the client catches become your emergency.

Downstream, stale data drives bounces. Practitioners keep bounce under 2%; above 4% burns sender reputation faster than warmup can repair, because inbox providers correlate bounces with low-quality sending [4]. One stale list degrades every future campaign.

05

The B2B Data Quality Framework (five dimensions)

Scrapoint's proposed framework — a practical lens, not an official standard.

1. **Freshness** — is the record current / recently verified?
2. **Completeness** — are the fields you use populated?
3. **Consistency** — are values standardized?
4. **Uniqueness** — are duplicates removed?
5. **Relevance** — does the record still fit the ICP?

A database can score high on volume and low on all five.

06

Measuring Database Health

A simple internal sampling method (not an industry metric):

1. Pull a random sample of 100 records.
2. For each, check: current title, current company, location, employment,

profile URL valid, duplicate status, ICP fit.

3. Score **freshness** = fresh records / sampled records x 100.
4. Repeat per dimension for a rough health profile.

Run it quarterly on the same method and you have a trend line — the first time most agencies can see decay instead of guessing at it.

07

Data Refresh Strategy

Do not refresh everything equally. Prioritise by value x staleness:

- **High-value accounts** — refresh most often.
- **Active-campaign contacts** — refresh before each campaign.
- **Older records** — refresh on a longer cycle.
- **Priority segments** — follow volatility (senior roles, fast industries).

Re-collect current data keyed on the LinkedIn profile URL so refreshes update records in place rather than creating duplicates.

08

From Database to Data System

The shift that ends the hidden cost is conceptual: stop treating prospect data as a file you finish; run it as a system.



- **Collect** current data from the source.
- **Normalize** into one schema.
- **Store** in a searchable, deduplicated system of record.
- **Refresh** on a priority cadence.

Data cleaning is an event. Data freshness is a process.

09

The Cost-of-Decay Calculator

You do not have to rely on a headline figure. You can estimate your own exposure. The model keeps **measured labour** (what a timesheet can count) separate from **estimated campaign loss** (which you should report on its own, so the same bad records are never counted twice).

Annual data labour = (Research + QA + Rework) x 12
Research = (Records refreshed per month / Throughput per analyst) x Analyst cost
QA = QA headcount x QA cost
Rework = Research x Rework rate
Campaign loss (estimated, reported separately)
= Stale contacts x Cost per wasted touch

Variable	What it is	How to get it
Throughput per analyst	Records a researcher verifies per month	Your own time study (~3,000 is a common benchmark) [5]
Analyst / QA cost	Fully-loaded monthly cost per person	Payroll plus overhead
Rework rate	Share of records re-done after QA finds an error	Your QA logs
Stale contacts	Records wrong on a key field	A sample audit (see the health check)

Keep the two totals apart. Labour is what decay costs you to fight; campaign loss is what it costs you when you lose. Adding them together counts the same bad records twice.

10

Worked Example: A Mid-Size Agency

Illustrative example. Every figure below is a stated assumption for demonstration — not a measured result, an industry average, or a Scrapoint benchmark. Replace the assumptions with your own numbers.

A hypothetical agency maintains 100,000 prospect records and refreshes or builds 25,000 a month. Assume an analyst verifies 3,000 records a month at a fully-loaded \$2,500/month, one QA reviewer per four analysts at \$3,000/month, and a 12% rework

rate.

Line item	Assumption	Monthly
Research	25,000 / 3,000 = ~8.3 analysts x \$2,500	~\$20,800
QA	2 reviewers x \$3,000	\$6,000
Rework	12% of research	~\$2,500
Measured labour subtotal		~\$29,300
Annualised	x 12	~\$352,000

Campaign loss sits on top of this and is estimated separately — bounced sends, damaged sender reputation, and reps working dead contacts. The point is not the exact total; it is that the number is large, countable, and recurring — and that most of it is spent fighting decay rather than finding new prospects.

11

A 30 / 60 / 90-Day Plan

A one-time cleanup cannot fix a continuously-decaying asset. A realistic rollout treats freshness as a process you stand up, not a project you finish.

Window	Focus	Key actions	Outcome
Days 0–30	Measure	Sample-audit 100 records; score freshness; estimate labour with the calculator	A baseline number to manage
Days 31–60	Prioritise	Segment by value x staleness; set a refresh cadence per segment; pick a stable key (profile URL)	A priority-ranked refresh plan
Days 61–90	Operate	Re-collect current data for the highest-value, oldest records in place; move QA to spot-checks; track freshness monthly	Freshness maintained, not restored

12 Where Scrapoint Fits

Scrapoint is the collection + normalization + storage + refresh layer of that system — factually:

- Collects public **LinkedIn and Sales Navigator** data (profiles, companies, lead/account searches).

- Normalizes into **one schema**, deduplicated on the profile URL.
- Stores records in a searchable **Data Vault**; exports to CSV / JSON / Excel.
- Re-collects for **refresh**, keyed on a stable identifier.

What it does not do: find personal emails or phone numbers, or invent data a person never published. It keeps your database in sync with what LinkedIn currently shows — continuously, at a scale manual teams cannot match.

13 Next Step

If your team spends significant time collecting and refreshing LinkedIn data, see how Scrapoint can own the collection layer so your people can focus on judgment and client outcomes.

Book a demo · scrapoint.com/product · Start free with 1,000 credits

Sources

1. Gartner — poor data quality costs organizations an average of \$12.9M/year (via Data Axle): <https://www.data-axle.com/resources/blog/stop-letting-dirty-data-drain-your-business/>
2. Landbase & Aircscale — B2B contact-data decay rates: <https://landbase.com/blog/data-decay-rate-statistics> · <https://airscale.io/blog/b2b-contact-data-decays-faster-than-you-think-here-is-the-evidence>
3. Second Talent — employee retention / tenure statistics 2026: <https://www.secondtalent.com/resources/employee-retention-statistics/>
4. Smartlead — cold email bounce rate & sender reputation: <https://www.smartlead.ai/blog/cold-email-bounce-rate>
5. Leadspace — SDR account-research time: [https://www.leadspace.com/resources/why-sdrs-lose-hours-a-week-to-account-research-\(and-how-to-get-it-back\)](https://www.leadspace.com/resources/why-sdrs-lose-hours-a-week-to-account-research-(and-how-to-get-it-back))