# Chain of Custody

## The Bad Marine Crash-Record Pipeline

How a crash report becomes a record a state can stand behind - from the officer's form or the driver's phone, through a person's review, to the federal submission - with every handoff documented and every change signed for.

Version: October 3, 2026
Bad Marine LLC - Crash safety engineering, sized for agencies that don't have a seven-figure IT budget.

# Contents

1. **The short version**
   The whole journey on one page
2. **Where a report starts**
   Officer XML, scanned paper, keyed by hand, a driver's own report
3. **Paper Image Extraction**
   Turning a scan into data, on the agency's own hardware
4. **The Refinery**
   Intake, an AI that only suggests, a person who confirms, the Vault
5. **Crash Location Resolution**
   A confirmed point becomes an official location and roadway data
6. **The finished record**
   The Vault, keeping older systems fed, the Public DB
7. **Crosswalk**
   From the state's record to MMUCC6
8. **FARS Manifest**
   Fatal crashes and the FARS-only elements
9. **Putting the data to work**
   TTAP: asking the data questions; CADRE: where safety money goes
10. **Why separate products**
    Every handoff, and what crosses it
11. **Setting it up**
    What's configured for each state, the link to NHTSA, and the known limits
12. **Where to go next**
    Manuals and online demos

# 1. The short version

A crash report starts as one person's account of what happened - an officer's or a driver's own - and has to be read, checked, corrected, placed on the map, filed as the state's official record and translated into the national format NHTSA relies on before it counts as clean, federally compliant data.

Most vendors sell that whole journey as one platform you adopt all at once. Bad Marine builds it as **separate products, each doing one real job**, joined only by a documented handoff. An agency can run the whole chain, or add only the piece it's missing.

The name says what it's built around: at every link the record is handed over in a documented form, a person signs off on what the software proposed, and every change is kept.

Downstream (chapter 9): **TTAP** for analysts' questions, **CADRE** for benefit-cost cases.

| | |
|---|---|
| **COMES IN** | **Officer XML · scanned paper · keyed by hand · a driver's own report**<br>Electronic reports from officers' reporting systems, and paper - including reports drivers file from their phones with **Mobile Driver Report**. |
| ▼ *scanned paper only* | |
| **READ** | **Paper Image Extraction**<br>Reads every box on a scan - typed, handwritten and checked - with a confidence and a location for each. Runs on the agency's own hardware. |
| ▼ | |
| **CLEANED & CONFIRMED** | **The Refinery**<br>Files each report into a case. A local AI suggests corrections; a person confirms every value. A second team places the crash on the map at the same time. |
| ▼ *alongside, in the map lane* | |
| **LOCATED** | **Crash Location Resolution**<br>An official location, plus the roadway data for that spot. |
| ▼ | |
| **FINISHED** | **The Vault**<br>Only reports that passed both lanes and every blocking rule. |
| ▼ | |
| **NATIONAL FORMAT** | **Crosswalk**<br>Maps the record onto MMUCC6 and checks the federal edit rules. |
| ▼ *fatal crashes also go to* | |
| **FATAL CRASHES** | **FARS Manifest**<br>The FARS analyst codes the FARS-only elements with the whole case file on screen, and sends the case to NHTSA. |
| ▼ *the finished data is also read by* | |
| **ANALYSIS** | **TTAP**<br>Analysts ask the finished data questions and check every number against its records. |

# 2. Where a report starts

A report enters the pipeline in one of four ways. They start differently, but they all end up in the same review.

| Way in | What arrives | First stop |
|---|---|---|
| **Officer's electronic report** | Structured XML sent straight from the officer's records-management or field-reporting system. Most reports arrive this way. | The Refinery's intake. |
| **Scanned paper** | A paper report - from an officer, or mailed or faxed in by a driver - scanned to a PDF. | Paper Image Extraction, then the Refinery. |
| **Keyed by hand** | A scan the agency would rather key than read automatically, or one that can't be read. | The Refinery's Manual entry screen. |
| **A driver's own report** | A driver who wasn't investigated by an officer files their own report from a phone with **Mobile Driver Report**. | The state's report inbox, then the Refinery. |

## 2.1 Mobile Driver Report - the driver's side

Many crashes are never investigated by an officer, and the driver is required to file a report of their own - usually on a paper form. **Mobile Driver Report** is a free public web app that walks a driver through the state's own form one screen at a time, on a phone, in English or Spanish. At the end it fills in the official form as a PDF and, if the driver chooses, a structured data file the state can load electronically, and hands both to the phone's share sheet to email in.

- **Nothing touches a Bad Marine server.** The form, the diagram, the PDF and the data file are all built in the driver's own browser; the draft is erased once the report is sent.
- **The data file skips the typing.** A report that arrives with its data file can be loaded electronically instead of being keyed from paper.
- **The crash diagram** is drawn on the phone as part of the report.

> **Why it matters to the agency.** Every driver report that arrives as data instead of handwriting is one less page to scan, read and key.

# 3. Paper Image Extraction

A scanned report is just a picture: nothing in it can be searched or checked until someone reads it. **Paper Image Extraction** (ppie) does that reading. For every scan it:

- recognizes the form from its own printed title - nobody has to say which form it is;
- straightens the page and lines it up against a blank copy of the same form, even when it was fed in crooked;
- subtracts the printed form, so only what the officer or driver wrote is left;
- reads every box - typed and handwritten text, code boxes and check marks;
- reports each reading with a **confidence** (0 to 1) and **exactly where on the page** it was found.

The result is one structured envelope per report, sent to the crash system along with the scan. In the Refinery, a reviewer clicks a value and the page jumps to the box ppie read it from.

> **It reads; it doesn't decide.** If the box says 04, ppie reports 04. Turning 04 into "Cloudy", filling in "Not Stated" or checking the report against the rules is the crash system's job, where a person reviews it.

## 3.1 On the agency's own hardware

ppie runs on the agency's own machine with no internet connection - the reading engine is built in and nothing is looked up online. The scans never leave it except to go to the agency's own crash system. Records staff don't operate it at all: scans dropped in a watched folder are read and sent on automatically. New forms are taught to it on a training screen, without programming.

Electronic reports skip this step entirely - they already arrive as data.

# 4. The Refinery

Every report - electronic, read by ppie, keyed by hand or filed by a driver - lands in the **Refinery**. It is the layer before a state's crash database: when the Refinery is done with a report, every value has been checked by a person and the crash has been placed on the map.

## 4.1 Cases, not loose reports

Each report is filed into a **case** - every report about one crash in one place. When an amendment, a supplement or a driver's report about the same crash arrives later, it files into the same case and the reviewer works only what changed. A report the Refinery can't match on its own is held for a person to decide: this case, a new case, or management.

## 4.2 An AI that only suggests

Before anyone opens a report, a local AI model reads it, checks it against the state's own rules and trusted federal lookups (such as the NHTSA VIN decoder and the FMCSA carrier registry), and attaches a proposed value and a confidence score to the fields it would change. **It never changes a value on its own.**

> **Air-gapped.** The AI model runs entirely on the agency's own server, from a file, with no network connection of any kind. Crash data is never sent to an outside AI service.

## 4.3 A person confirms every value

A reviewer works each report **once, in place**: the officer's value, the AI's suggestion and its confidence sit side by side, with the report image beside them. The reviewer accepts or corrects each one and marks the data complete. There is no send-back loop - a report the reviewer can't fix goes to management, is answered, and comes back or is resolved there.

That confirmation isn't optional and can't be skipped. It is the safeguard that keeps a wrong guess - the officer's or the AI's - from becoming the official record.

## 4.4 Two lanes at once

| Lane | Who | What they do |
|---|---|---|
| **Validation** (data) | Reviewers | Confirm the report's data field by field against the report image. |
| **Location** (map) | Mappers | Place the crash on the map, snap it to the road network, and fill in the official location and roadway data through Crash Location Resolution (chapter 5). |

When both lanes are done and no blocking rule is broken, the report goes to the **Vault** automatically (chapter 6).

## 4.5 Everything around the report

- **Specialist work:** work zone review, and a FARS screen for fatal crashes - or FARS Manifest, if the state uses it (chapter 8).
- **Search for the front desk:** find a crash report for a customer and send its documents.

- **Supervision:** team dashboards, operations metrics and downloadable reports.
- **Documents:** the report PDF, the original XML, photos and later documents are kept with the case and handed to the state's imaging system, the long-term home for documents.
- **Getting better over time:** every correction a person makes is kept. A Learning screen finds fields where people keep overriding the AI and lets an administrator turn the pattern into a correction the AI then suggests - so there is less to fix each month.

# 5. Crash Location Resolution

A latitude and longitude isn't a location a state can file. **Crash Location Resolution** is a web service that takes a confirmed crash point and works out, against the state's own GIS data, in one answer:

- **on a state route:** the route, reference post and log mile, calibrated between the surveyed mile markers on either side of the crash;
- **on a local road:** the nearest street, intersection, bridge or railroad crossing, the cross street, distance and direction, and a written narrative;
- bridge context, and the distance to the nearest town when the crash is outside city limits;
- the **roadway (MIRE) data** for that spot - speed limit, traffic volume, ownership, district, rural or urban and more - each either filled in or marked unavailable, with the reason.

Where the state's GIS can supply a roadway value, that value wins over what was reported at the scene: the officer's report is not the source for roadway data. The service resolves exactly the point it's given and never moves it - placing the point is the reviewer's call. It has been measured against thousands of a state's own finalized crash locations.

## 5.1 In the pipeline, and on its own

In the Refinery, a mapper confirms or moves the point and clicks **Map it**; the answer fills in the location fields of the same record the data reviewers are working. If the point moves again, the old answer is marked out of date until it's mapped again.

The service stores nothing and needs nothing else from Bad Marine. Any system that has to turn a coordinate into a standard location reference can call it directly - one crash at a time, or a whole CSV of points in one batch.

# 6. The finished record

When both the data and the location work are done and the report passes every blocking rule, it is written to the **Vault** - the state's finished, validated record, on the agency's own servers. At that point the report is done as far as day-to-day operations are concerned: clean, confirmed and usable.

> **A validity gate, not a lock.** A report in the Vault can still be pulled back through management for more work and put back. The Vault always holds the latest approved version.

## 6.1 The three stores

| Store | What it holds | Who sees it |
|---|---|---|
| **Public DB** | Each report's data as it arrived, minus the fields the state leaves out (by default every name, address, birth date, phone, licence, plate, VIN, policy and citation number, and the narrative). Optional. | The one store meant for outside eyes. |
| **Hub** | The working copy: the AI's suggestions, every edit, the case, the conversation with management, the full activity log. | Refinery staff, by role. |
| **Vault** | The finished record of every report that passed both lanes and every blocking rule. | Refinery staff; read by Crosswalk and Legacy output. |

## 6.2 Nothing has to be replaced

A state rarely retires its older crash systems on a schedule. **Legacy output** maps Refinery data into the state's existing tables and keeps them fed - live as reports finish, on a schedule, or as a one-off re-push - so the reports, queries and partner feeds built on those tables keep working.

# 7. Crosswalk

A state's crash record is written in the state's own format; NHTSA wants **MMUCC 6th Edition** (the Model Minimum Uniform Crash Criteria). **Crosswalk** picks the record up where the state's system leaves it and:

- maps every crash, vehicle and person field onto its MMUCC6 element, with a code-by-code crosswalk wherever the state's codes differ from the federal ones;
- checks the mapped result against **130 of the 131 MMUCC6 edit rules**, and lists every violation by name instead of dropping it or passing it through quietly;
- keeps the state's own values that have no MMUCC6 home in a **sidecar** table, so nothing the state collects is lost on the way to the federal format;
- lets the state choose which personal fields are left out of a transmission.

Crosswalk never touches how an officer files a report and doesn't require the state to rebuild anything: it works from whatever the state's crash system already produces. With the Refinery in place, it reads the Vault. It has been built and tested on real public crash data from five jurisdictions with very different formats.

## 7.1 People still sign off

A wrong code crosswalk would quietly corrupt federally reported data, so Crosswalk follows the same rule as the Refinery: software proposes a mapping, a person reviews it, and only then is it locked. Rule violations are worked the same way - each one is fixed, excluded or allowed by a person, and the decision is kept.

# 8. FARS Manifest

Every fatal crash also goes to FARS, NHTSA's Fatality Analysis Reporting System - and FARS needs elements a state crash report never captures. **FARS Manifest** is the FARS analyst's workspace for them. With the officer's report, the state's record in plain language, and every other document in the case (supplements, driver reports, amendments, lab reports) on screen, the analyst:

- logs a **pre-report** the moment they hear of a fatal, and matches it to the officer's report when it arrives;
- codes the **53 FARS-only elements** on ready-made forms with NHTSA's official code lists;
- keeps each person's **Alcohol Test** current as lab results arrive;
- clicks **Send to NHTSA**, which packages the state's record with the FARS coding - and **Retransmit** when a late lab result changes the answer.

FARS Manifest is not a stage of the pipeline; it works alongside it. It reads finished fatal cases from the state's crash system - the Refinery, or another system speaking the same documented contract - which stays the system of record. FARS Manifest writes back only what the analyst does: that a case was sent, the FARS coding, and alcohol updates.

# 9. Putting the data to work

Clean, located crash data is the raw material for two kinds of question: **what is happening on our roads**, and **where do we spend safety money, and will it work?**

## 9.1 TTAP - asking the data questions

**TTAP** (Traffic & Transportation Analysis Probing) is where a state's crash data analysts ask questions of the finished data. The analyst builds the question on screen as one readable sentence - for example, count crashes from officer reports where the crash was fatal and a vehicle was a hit and run, grouped by day of week - down to the vehicle and the person, and gets a table, a chart, or the list of crashes behind any number. Clicking a row opens the whole report. Nobody writes code, and nothing a person types becomes SQL.

TTAP only reads. It takes a read-only feed from the Refinery into its own copy, refreshed on a schedule, and never changes a crash report or opens the Refinery's databases. Names, addresses, plates and the narrative are not sent to it. Every result says how fresh the data is, and stale data, missing sources and crashes not yet FARS-coded are stated on screen rather than silently left out. Where the state has set it up, an analyst can type a question in plain words and the Refinery's local AI model fills in the query builder - a person still reads it and presses Run.

## 9.2 CADRE - where safety money goes

**CADRE** (Crash Analysis & Decision-support for Roadway Evaluation) turns an agency's own crash history into a defensible benefit-cost case for a safety project - a roundabout, a signal, rumble strips, a road diet. It finds the crashes at a site or along a corridor, applies countermeasures from the FHWA CMF Clearinghouse, adjusts for regression to the mean with the Highway Safety Manual's Empirical Bayes method, and produces a benefit-cost ratio with a Monte Carlo range and a report a grant reviewer can follow, every formula traced to its federal source. It also screens a whole network for sites worth a closer look.

CADRE is a licensed Windows product. It loads crash history from a workbook export and roadway inventory from CSV, so it works whether or not the rest of the pipeline is in place - but crash data that has been through the Refinery and Crash Location Resolution is exactly the located, confirmed history its math depends on.

## 10. Why separate products

Each product here is its own codebase, with its own runtime and its own release schedule. The only thing connecting any two of them is a documented handoff at that one boundary:

| From | To | What crosses the boundary |
|---|---|---|
| Mobile Driver Report | The state | An email: the official form as a PDF, plus an optional structured data file. |
| Paper Image Extraction | The Refinery | One documented JSON envelope per report (every reading, its confidence and its place on the page), plus the scan. |
| The Refinery | Crash Location Resolution | One web call per confirmed point; one answer back. Nothing is stored on the service. |
| The Refinery (Vault) | Crosswalk | A read of the finished record. |
| The Refinery | FARS Manifest | A documented HTTP contract: fatal cases out; send status, FARS coding and alcohol updates back. |
| The Refinery | Older state systems | Legacy output into the state's existing tables. |
| The Refinery | TTAP | A read-only feed with a key: what the state has, its fields and code lists, and what changed. Names, addresses, plates and narratives are not sent. |
| Any crash system | CADRE | A crash-data workbook and roadway inventory CSVs. |

So an agency that already runs its own paper intake, its own records system or its own MMUCC reporting can add exactly the piece it's missing without touching what already works - and can rely on any one piece without ever adopting the rest. That's checked, not just claimed: the Refinery, for example, builds and passes its full test suite on a machine with no other Bad Marine code present at all.

# 11. Setting it up

Each product is installed against a state's own systems. These are the normal setup steps, done once, not shortcomings.

## Set up at installation

| Item | How it's done |
|---|---|
| **Sign-in** | The Refinery, FARS Manifest and TTAP use the state's own directory sign-in, so staff sign in with their existing state account. The connection to the state's directory is configured during installation. |
| **State connections** | The imaging system, driver-licence lookups and older-system feeds are configured and tested against each state's own systems at installation. |
| **New forms and formats** | A new paper form is taught to ppie once on its training screen, and a new electronic report format or state is mapped once by an administrator. After that its reports are read like any other. |
| **Which pieces to adopt** | Each product stands alone. An agency without ppie keys scanned reports in the Refinery; one without Crash Location Resolution fills the location fields in by hand. |

## The link to NHTSA

The Refinery and Crosswalk produce the MMUCC6 export, and FARS Manifest's Send to NHTSA packages each fatal case in NHTSA's own XML schema. NHTSA publishes that schema; connecting submissions into NHTSA's systems is NHTSA's side of the link, and NHTSA is the one that sets it up.

## Known limits

| Item | Where it stands |
|---|---|
| **MMUCC6 edit rules** | Crosswalk checks 130 of MMUCC6's 131 edit rules. The remaining one - that a crash's coordinates fall inside its county - needs county boundary data supplied by the state. Rules that compare the two vehicles in the first harmful event are checked on two-vehicle crashes, where that pair is certain. |
| **GIS coverage** | Crash Location Resolution is only as complete as the state's GIS layers; where a layer has no value for a spot, the answer says so instead of guessing. |

# 12. Where to go next

Every product here can be tried online with made-up data and no sign-up. User manuals are free to download from **badmarinellc.com**; technical manuals are shared on request.

| Product | Manual | Try it |
|---|---|---|
| Paper Image Extraction | User Manual | Online demo with a sample scan |
| The Refinery | User Manual | Online demo - pick a role and work the queues; resets nightly |
| Crash Location Resolution | User Manual for System Integrators | Online evaluation page |
| Crosswalk | User Manual | Online demo |
| FARS Manifest | User Manual for FARS Analysts | Online demo |
| TTAP | User Manual for Crash Data Analysts | Online demo - the Query Studio on made-up crashes |
| CADRE | User Guide; Technical Manual on request | Online demo; licensed for agency use |
| Mobile Driver Report | User Manual for Drivers | Live, free to the public |

Which piece do you need? Some agencies need one stage; others want to plan the whole chain. Either conversation is a fit.

**Bad Marine LLC** · 200 S 21st St Ste 400a, Lincoln, NE 68510 · **sean@badmarinellc.com** · **badmarinellc.com**