

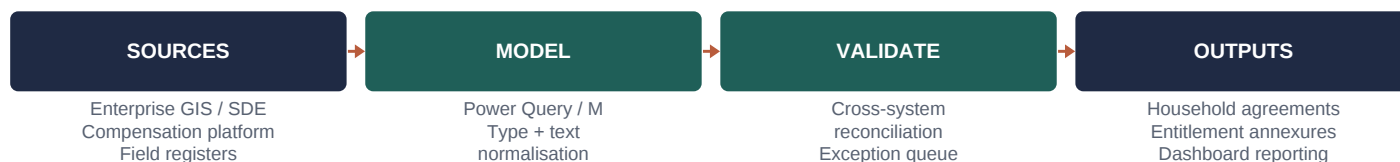
Architecture of an Auditable Record

Resettlement data environment, Mozambique LNG · 2013–2025

Eight thousand affected households. Twelve years. One governing constraint: every number in every report had to lead back to a household-level source record that an external auditor could reach, under IFC Performance Standard 5. That constraint is what produced the architecture below. Reconciliation, exception handling and handover documentation are not additions to it — they are the reason it has the shape it has.

8,000+ affected households in the data environment	12 years of continuous operational stewardship	2 languages, in every generated document	5 source systems reconciled each cycle
---	---	---	---

The pipeline



What the validation layer actually tested

Not schema conformance. The checks below target the specific ways a compensation record goes wrong in the field, where a status flag and the evidence for that status are captured by different people at different times.

Check	What it compares	The failure it catches
Agreement signature	Signed flag against the presence of a signature date	A household marked as agreed with no dated agreement behind it
Finance submission	Submitted flag against the submission date	Payments reported as in progress that were never lodged
Proof of payment	Paid flag against the proof-of-payment date	Closure claimed without the receipt that proves it
Identity duplication	Agreement identifiers grouped and counted	One household entering the register twice under variant codes
Amount duplication	Identifier and amount together	A compensation amount paid, and paid again
System reconciliation	Tracker against the enterprise GIS, full outer join	Records present in one system and absent from the other, in both directions

Each check returns a three-state result: consistent, inconsistent, or **Check** — the explicit third state for a record whose flag is neither yes nor no. That third state is the important one. A two-state check silently absorbs the cases nobody has decided yet; a three-state check makes the undecided population countable and puts it in front of a human.

The reconciliation, in the transformation language itself

A representative excerpt, paraphrased and stripped of paths and identifiers. The pattern is a full outer join rather than an inner one, so that unmatched records on *either* side survive into the output instead of disappearing from the comparison.

```
// Compare the compensation tracker against the enterprise GIS extract.
// Full outer join: absence on either side is itself a finding.
let
  Source          = CompensationTracker,
  Renamed         = Table.RenameColumns(Source, {"Ag", "Comp.Ag"}),
  Joined          = Table.NestedJoin(Renamed, {"Comp.Ag"},
    GisAgSummary, {"Ag"}, "Data", JoinKind.FullOuter),
  Expanded        = Table.ExpandTableColumn(Joined, "Data",
    {"Ag", "CCN", "Phase", "Amount_OR", "Amount_NR"}),
  Zeroed          = Table.ReplaceValue(Expanded, null, 0,
    Replacer.ReplaceValue, {"Amount_OR", "Amount_NR", "Final"}),
  Grouped         = Table.Group(Zeroed, {"Data.Ag", "Comp.Ag", "Final"},
    {"Amount_OR", each List.Sum([Amount_OR]), type number})
in
  Grouped

// Three-state consistency check: flag against the evidence for the flag.
AddedCheck = Table.AddColumn(Source, "HHA_Check", each
  if [HHA_Signed] = "Yes" then [HHA_Date] <> null
  else if [HHA_Signed] = "No" then [HHA_Date] = null
  else "Check")
```

Generated documents, not assembled ones

Household compensation agreements and entitlement annexures were generated from the validated model rather than typed from it. Asset classes, quantities, units, rates and amounts were grouped and rendered into bilingual output with the Portuguese asset vocabulary held as data, not as translation performed per document. The consequence that mattered operationally: a correction to a household record propagated to every document derived from it, and no document could contain a figure the model had not validated.

Handover

The system was documented and handed over as an operating procedure, not as a file: refresh sequence, dependency order, the errors an operator should expect to see and what each one means, and how to verify the outputs are correct before releasing them. A data environment that only its author can run is a single point of failure wearing the costume of an asset.

Client: Anadarko, subsequently TotalEnergies. Delivered remotely across distributed multinational teams and time zones. This document contains no client file paths, account identifiers, or household-level data; code excerpts are paraphrased and anonymised.

Liezl Coetzee · Sociable Systems · sociable.systems · liezlc@gmail.com