



Legacy System Data Migration and Data Quality

PROBLEM STATEMENT

RD's legacy mainframe systems (PLAS, DLS, CLSS, GLS, LoanServ, AMAS, ECMS) carry decades of data with inconsistent field definitions, partial documentation, and aging extraction interfaces, and any cohort, modification, or audit work depends on extracts that must be rationalized before they are usable.

WHY IT MATTERS

Bad data is the upstream cause of most cohort-model staleness, ULO certification gaps, reconciliation breaks, and audit findings. Investing in legacy data quality compounds across every other workstream.

HSG'S APPROACH

1. Stand up a data-quality program with profiling, definition-mapping, lineage documentation, and cleansing rules across each legacy system extract.
2. Build a canonical data model with field definitions tied to USSGL, OMB DAIMS, and FCRA cohort conventions.
3. Use Python / dbt / DuckDB / SQL Server to build an automated data-quality pipeline that surfaces and tracks data-quality exceptions over time.
4. Produce a data-quality scorecard per legacy system with monthly trending.
5. Document everything in a data dictionary suitable for the upcoming USDA Loan Modernization migration.

EXPECTED DELIVERABLES

- Data-quality profiling report per legacy system
- Canonical data model and data dictionary
- Automated data-quality pipeline
- Monthly data-quality scorecard
- Cleansing-rule library with lineage documentation

EXPECTED OUTCOME

Reduce data-quality exception count by 50% within 12 months and produce a migration-ready data dictionary in advance of the loan-modernization cutover.

ENGAGEMENT TEAM

PRIMARY LEAD

Michael Nelson

Senior Financial Analyst · HUD OAS Bench

CONTRIBUTORS

Brad Dillman

Senior Quantitative Modeler · Federal Credit-Reform Modeling Lead

Jelani House

Engagement Manager · Business Development Lead · AI Lead · FCRA & Budget Execution Lead