International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Standard Operating Procedures for Governed Master Data Consolidation in EBX5

Kinshuk Dutta

Submitted:03/05/2018 **Revised**: 18/06/2018 **Accepted**: 27/06/2018

Abstract- Master data consolidation programs often fail not because the target platform lacks technical capability, but because operating controls, role boundaries, stewardship procedures, and exception-handling patterns are not formalized early enough. This paper presents a standard operating procedure (SOP) framework for governed data consolidation using EBX5. The framework defines a repeatable operating model for receiving data from source systems, validating and loading records into a landing zone, transforming and standardizing records through staging, applying matching and survivorship rules, resolving ambiguous clusters through stewardship, and publishing governed golden records. The paper emphasizes the separation of automated system tasks from accountable user tasks, the importance of stewardship role design, and the controlled onboarding of new data providers. The result is a practical governance-oriented architecture for reducing implementation variability and improving repeatability in multi-domain master data management deployments.

Index Terms- *EBX5, master data management, standard operating procedure, data consolidation, data governance, data stewardship, golden record, survivorship, data quality, source-system onboarding.*

I. Introduction

Enterprise master data management (MDM) initiatives require more than a repository and a matching engine. They require a disciplined operating model that determines how data enters the environment, how source defects are handled, how standardized records are prepared, how matching decisions are governed, and how a reliable golden record is made available for enterprise consumption. Without this operating model, even technically correct implementations can become brittle, inconsistent, and difficult to scale across domains or business units.

This paper presents a standard operating procedure framework for governed data consolidation using EBX5. The framework is designed for implementations that follow a landing, staging, and mastering pattern, where external source data is received, inspected, normalized, matched, stewarded, and mastered under explicit governance controls. Although the examples are expressed in EBX5 terminology, the principles apply broadly to multi-domain MDM programs that must balance automation with human accountability.

The proposed SOP framework is motivated by four recurring implementation challenges. First, source data is often loaded before clear acceptance and rejection rules are defined. Second, staging logic is frequently treated as an implementation detail rather than a governed transformation boundary. Third, stewardship responsibilities are assigned too late, causing ambiguous match clusters to accumulate without clear ownership. Fourth, new source systems are connected without sufficient profiling, crosswalk analysis, or validation of survivorship impact. A formal SOP addresses these risks by defining the activities, roles, decisions, and controls required for repeatable data consolidation.

II. Background And Problem Statement

Data consolidation is the process of collecting records from multiple source systems and producing a trusted representation that can be used by enterprise processes. In a master data context, consolidation includes technical movement of data, structural normalization, quality assessment, matching, survivorship, stewardship, and

Principal Solutions Engineer
Orchestra Networks
New York, USA
Email: kinshuk.dutta@orchestranetworks.com; dutta.kinshuk@gmail.com
Corresponding Author: Kinshuk Dutta

publication. The difficulty lies in the fact that these activities cut across both IT and business accountability. The system can detect duplicates or quality exceptions, but business owners and stewards must often determine whether a record should be accepted, corrected, rejected, merged, or retained as a distinct entity.

A platform-centric implementation approach may define tables, import jobs, and match rules, but still fail to specify what should happen when a source feed contains invalid values, when a candidate duplicate falls into an uncertain score range, or when a new provider introduces fields that conflict with established semantics. Therefore, the SOP must be considered part of the technical architecture, not merely a training artifact.

III. SOP Design Principles

The SOP model proposed in this paper is based on six design principles. First, each zone in the consolidation pipeline must have a clear operational purpose. Landing zones preserve the source-oriented view of incoming data, staging zones prepare data for mastering, and mastering zones store governed records in the structure required by enterprise usage. Second, every automated step must have a corresponding exception path. Third, every exception path must have an accountable role. Fourth, stewardship must be treated as a designed control point rather than an ad hoc clean-up activity. Fifth, source onboarding must be controlled through profiling, validation, and governance approval. Sixth, the model must preserve traceability from source values to mastered output wherever possible.

These principles allow an implementation team to distinguish between technical success and governed success. A feed is not successful merely because records are imported. It is successful when the data has passed the required controls, quality exceptions are handled consistently, identity ambiguity is resolved appropriately, and the resulting mastered values are fit for downstream use.

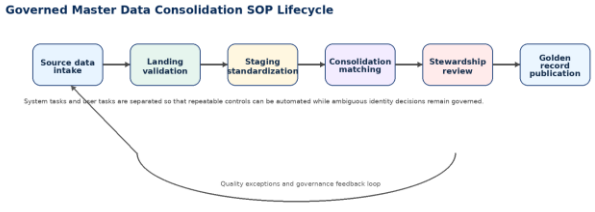

Fig. 1. Governed SOP lifecycle for EBX5 data consolidation.

IV. Operating Roles

A consolidation SOP must define operational roles before the first production feed is loaded. The principal roles include the architect, database administrator, system administrator, solution developer, business analyst, business user, data steward, and data steward manager. These roles separate platform operation, solution design, business interpretation, and stewardship decision-making. The distinction is important because consolidation problems often arise when a technical team is forced to make business identity decisions, or when business stewards are asked to resolve issues without visibility into the transformation and matching process that created them.

The data steward role is especially central. A data steward manages information quality for one or more subject areas and is responsible for investigating duplicates, validating values, resolving identity issues, and maintaining business rules. Depending on the organizational model, stewards may be assigned by data object, business function, process, system, or matching responsibility. A mature consolidation implementation should identify these stewardship categories explicitly and align them with workflow routing, permissions, dashboards, and service-level expectations.

Operating Roles in a Governed EBX Consolidation Program

Role	Primary responsibility
Architect	Defines solution pattern, integration topology, and data model principles.
DBA / System Admin	Maintains platform reliability, database availability, backups, and operations.
Solution Developer	Implements configurations, services, workflows, and controlled customizations.
Business Analyst	Translates business needs into data rules, process logic, and acceptance criteria.
Data Steward	Investigates quality exceptions, resolves matching ambiguity, and validates survivorship.
Steward Manager	Monitors quality metrics, escalations, workload, and operating discipline.

Fig. 2. Core operating roles in a governed EBX5 consolidation program.

V. Data Loading Operation

The data loading operation begins after external data is received and placed into the environment. The loading operation can be manual or automated, but in both cases the SOP must define what constitutes an acceptable feed. In an automated model, a system task creates a controlled draft area in the landing zone and loads the provider file as received. The system then performs validation checks against structural, referential, and business-rule constraints. If errors are identified, the process routes an exception to the appropriate process data steward or operational owner.

The steward then determines whether the data should be corrected inside the governed environment, rejected and returned to the provider, or accepted with documented exceptions. After the user decision is recorded, the draft area can be merged into the prime landing zone or archived. This

distinction between draft and prime areas is critical because it prevents unreviewed source data from becoming part of the controlled consolidation flow.

The loading operation should produce a minimum audit record consisting of feed identifier, provider, load time, record count, validation outcome, exception count, responsible steward, decision, and final status. These fields allow operational teams to detect recurring provider issues and measure whether data quality is improving over time.

VI. Data Staging Operation

The staging operation transforms accepted landing data into structures and values suitable for mastering. Staging is not merely a temporary storage step. It is the point at which source-oriented representations are conformed to the enterprise model. Typical staging activities include value translation, data type normalization, reference-data standardization, address or communication decomposition, enrichment, validation, and preparation for matching.

The SOP should specify whether transformation failures are corrected in EBX5, returned to the source provider, or held in an exception queue. For example, a source field containing a date of birth as a string may need to be converted into a date type. A source prefix value may need to be standardized through cross-reference lookup. Address values may need to be split into structured components. Each transformation should be governed by a mapping rule, a validation rule, and an exception policy.

A common anti-pattern is to allow staging logic to evolve informally as scripts or one-off mappings. The recommended approach is to treat staging as a governed design layer with documented mappings, reusable transformation patterns, and stewardship visibility into exception cases. This makes the movement from landing to mastering auditable and repeatable.

VII. Data Consolidation Operation

The consolidation operation creates or updates the mastered representation of records. It includes matching, cluster creation, survivorship, golden record generation, and stewardship of ambiguous cases. Matching policies may be configured to detect similarity across names, addresses, identifiers, communication channels, or relationships. Depending on score thresholds, the system may auto-merge, route records to stewardship, or classify records as non-matches.

Survivorship rules determine which values should survive into the golden record. These rules may consider source trust, recency, completeness, manual steward selection, or explicit business preferences. The SOP must define how survivorship is applied, when stewards may override it, and how overrides are logged. In regulated or high-risk domains, the ability to explain why a value survived is as important as the value itself.

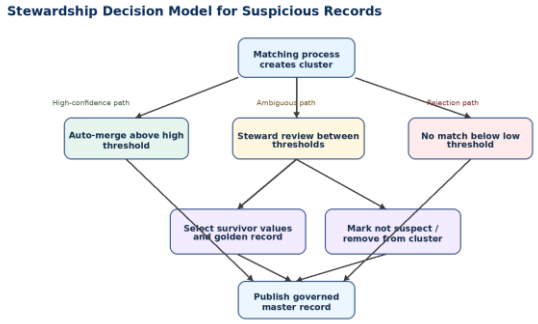

Fig. 3. Stewardship decision model for suspicious records and match clusters.

VIII. Matching Stewardship

Matching stewardship is required when the system identifies records as suspicious or when automated confidence is insufficient to resolve identity. The steward must be able to inspect clustered records, compare candidate values, identify incorrect groupings, manually remove records from a cluster, mark records as not suspect, select survivor values, and confirm the golden record. The SOP should specify the actions available to the steward and the conditions under which each action is appropriate.

A well-designed stewardship process prevents two forms of consolidation failure. The first is over-merge, where distinct entities are incorrectly combined. The second is under-merge, where duplicate entities remain fragmented across source systems. Both failures reduce trust. The stewardship process acts as a governance control to correct the system when statistical matching alone is insufficient.

The SOP should also define how recurring stewardship decisions influence future matching behavior. If a pair of records is repeatedly marked as not suspect, the system should avoid re-presenting the same pair without new evidence. If a specific source consistently causes ambiguous matches, provider quality or source-specific transformation logic should be reviewed. This turns stewardship from a manual queue into a feedback mechanism for continuous improvement.

IX. Modification To Accept New Sources

A consolidation platform must evolve as new source systems are added. However, onboarding a new provider can change the behavior of the entire mastering process. New fields may conflict with existing semantics, new identifiers may require crosswalk changes, and new data patterns may influence matching thresholds. The SOP therefore requires a controlled validation process before a new source is accepted into production consolidation.

The recommended onboarding process begins with provider request and sample data collection. The implementation team then creates or reviews a sample data model, imports sample records, profiles the sample data, evaluates candidate keys and distinctness, validates field-level quality, and performs linking analysis to understand how the new provider relates to existing mastered records. Only after these steps should the source be approved for production integration.

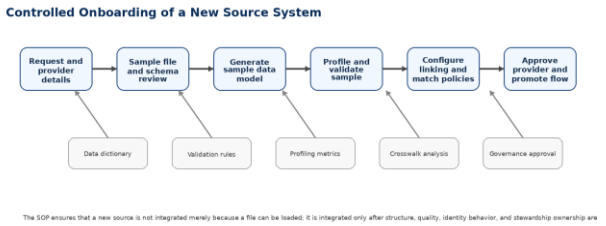

Fig. 4. Controlled onboarding process for a new data provider.

X. Controls And Metrics

The SOP should define measurable controls for the consolidation lifecycle. Loading controls include feed arrival, file format validity, record count reconciliation, and schema validation. Staging controls include transformation success rate, standardized value coverage, reference-data lookup failures, and validation exceptions. Consolidation controls include match rate, auto-merge rate, stewardship queue size, over-merge corrections, under-merge findings, and golden record publication latency.

These metrics support both operational monitoring and governance review. For example, an increase in stewardship queue volume may indicate that a provider changed its data format, that matching rules are too permissive, or that a reference data dependency has degraded. A high rate of manual override may indicate that survivorship rules no longer reflect business trust. By attaching metrics to each SOP step, the organization can diagnose systemic issues rather than treating exceptions as isolated incidents.

XI. Implementation Considerations

Several implementation considerations are important when applying this SOP framework in EBX5. First, permissions should reflect stewardship responsibility. Users responsible for matching review should have access to the records, clusters, and merge services required to perform that work, but not necessarily broad administrative privileges. Second, workflow design should distinguish between operational exceptions, data quality exceptions, and identity resolution exceptions. Each category may require different routing and escalation.

Third, the landing and staging models should preserve enough source detail to support traceability. Over-normalizing too early may destroy evidence needed for stewardship. Fourth, the mastering model should support survivorship and publication requirements rather than simply mirror the source structures. Fifth, the SOP should be tested during user acceptance testing, not introduced after go-live. If stewards cannot resolve real exceptions during test cycles, the operating model is not ready for production.

Finally, documentation should remain synchronized with configuration. SOP documents, workflow definitions, mapping rules, matching policies, and stewardship screens should be reviewed together. When the platform changes but the SOP does not, users lose trust. When the SOP changes but the platform does not, users are asked to follow a process that the system does not support.

XII. Discussion

The primary contribution of this paper is the treatment of SOP design as a first-class component of MDM architecture. Many data consolidation initiatives focus heavily on data models and match rules while underinvesting in operating discipline. The proposed framework shows that loading, staging, consolidation, stewardship, and source onboarding are not isolated workstreams; they are linked controls in a governed system.

The framework also demonstrates why human-in-the-loop design is necessary in master data consolidation. Automated matching can provide candidate clusters and confidence scores, but business identity remains contextual. A system may determine that two records are similar; a steward may determine whether that similarity represents the same real-

world entity. The SOP provides the procedural structure that allows this decision to be made consistently and auditable.

XIII. Limitations

This paper focuses on the operating model for consolidation rather than the mathematical design of specific matching algorithms. It also does not prescribe a single enterprise governance structure, since stewardship responsibility varies by organization, data domain, and regulatory environment. The framework should therefore be adapted to local governance policy, data sensitivity, domain complexity, and integration architecture.

XIV. Conclusion

A governed master data consolidation program requires a repeatable operating procedure that connects technical data movement with business accountability. The SOP framework presented in this paper defines how source data should be loaded, validated, staged, standardized, matched, stewarded, and mastered in EBX5. By separating automated system tasks from accountable user decisions, defining stewardship responsibilities, and formalizing the onboarding of new data providers, organizations can reduce implementation variability and improve trust in golden records. The result is a practical, scalable operating model for enterprise data consolidation in multi-domain MDM environments.

XV. Appendix A: Operational Checklist

Phase	Minimum SOP controls
Source intake	Provider identity, file/control totals, expected schema, received timestamp, security classification.
Landing	Draft area creation, validation check, exception routing, accept/reject/archive decision.
Staging	Transformation mapping, reference-data lookup, type conversion, quality exception handling.
Consolidation	Matching policy execution, threshold handling, cluster creation, survivorship application.
Stewardship	Review queue, cluster inspection, merge/no-merge decision, survivor value selection, audit trail.
Publication	Golden record status, publication timestamp, downstream notification, reconciliation.

XVI. References

[1] DAMA International, DAMA-DMBOK: Data Management Body of Knowledge, 2nd ed. Basking Ridge, NJ, USA: Technics Publications, 2017.

[2] C. Batini and M. Scannapieco, Data and Information Quality: Dimensions, Principles and Techniques. Cham, Switzerland: Springer, 2016.

[3] D. Loshin, Master Data Management. San Francisco, CA, USA: Morgan Kaufmann, 2008.

[4] A. Silberschatz, H. F. Korth, and S. Sudarshan, Database System Concepts, 6th ed. New York, NY, USA: McGraw-Hill, 2010.

[5] Orchestra Networks, EBX5 Implementation and Data Consolidation Field Materials, 2018.

[6] Orchestra Networks, EBX5 Standard Operating Procedure for Data Consolidation, 2018.

[7] TIBCO Software Inc., EBX and Data Quality Add-on Technical Documentation, 2018.

[8] TIBCO Software Inc., EBX Data Exchange Add-on Technical Documentation, 2018.