Matching and Survivorship Design Patterns for Party Data Consolidation in EBX5

Technical Research Article - October 2018

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

Abstract- Party data consolidation is one of the most difficult problems in multi-domain master data management because individual and organization records often arrive from heterogeneous source systems with inconsistent naming conventions, incomplete addresses, duplicated communication channels, and conflicting identifiers. This paper presents a set of practical design patterns for matching and survivorship in EBX5-based data consolidation programs. The patterns distinguish identity resolution from survivorship, separate single-table matching from relation-based matching, apply context-aware rules for person and organization records, and use threshold-governed stewardship to determine when automation is appropriate and when accountable human review is required. The paper also defines a survivorship matrix for preserving lineage while constructing trusted golden records. These patterns provide an implementation-oriented framework for reducing false merges, improving traceability, and making data stewardship decisions explainable in enterprise master data programs.

Index Terms- *Master Data Management, EBX5, data consolidation, data quality, matching, survivorship, golden record, stewardship, identity resolution, fuzzy matching, party data.*

I. INTRODUCTION

A master data management platform can centralize records and expose governed workflows, but it does not by itself solve the harder semantic problem of determining whether two source records represent the same real-world entity. In party data domains, this problem is intensified by name variation, address inconsistency, missing identifiers, legacy system conventions, and regional differences in data capture. A record for an individual may be expressed as a first name, last name, full name, nickname, and address combination. A record for an organization may be expressed through legal names, trade names, abbreviations, locations, and tax or registration identifiers. The design of matching and survivorship policies therefore becomes a core architectural responsibility, not an afterthought.

This paper presents design patterns for EBX5 party data consolidation. The goal is to provide a repeatable architecture for configuring matching policies, relation-aware rules, stewardship thresholds, and survivorship decisions so that the consolidation process remains explainable and operationally governable. The paper uses the term party to represent either a person or an organization and assumes that source-specific records must be consolidated into governed master records while retaining source lineage.

II. PROBLEM STATEMENT

The central challenge is that identity resolution and value survivorship are often conflated. Identity resolution asks whether two or more records describe the same entity. Survivorship asks which attribute values should represent the entity in the golden record. A system may correctly identify that records are probable duplicates but still require business judgment to determine whether the preferred address, communication channel, or classification should survive. Conversely, an aggressive matching rule may construct a false merge if it treats similar names or addresses as conclusive evidence of sameness without context.

The proposed framework addresses four recurring implementation risks: over-reliance on a single score, lack of separation between person and organization logic, loss of source lineage after consolidation, and insufficient stewardship controls for ambiguous clusters. These risks are common in multi-source MDM programs and can directly affect downstream analytics, compliance, customer engagement, and operational trust.

III. REFERENCE MATCHING PIPELINE

A robust party consolidation design begins with a pipeline that separates technical preparation from business decision-making. Source records are first standardized so that comparable attributes can be evaluated consistently. Candidate generation narrows the search space to records that could plausibly match. Matching policies then calculate similarity scores based on exact, fuzzy, and relation-aware rules. Clusters are created for records that meet policy conditions, and the resulting clusters are either automatically merged, rejected, or routed to stewardship depending on score thresholds and governance rules.

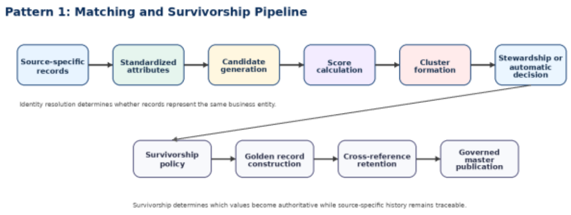

Fig. 1. Matching and survivorship pipeline for party data consolidation.

The pipeline in Fig. 1 emphasizes that matching is not the final step. After identity decisions are made, survivorship policies construct the golden record while preserving cross-reference mappings to source records. This design allows the consolidated record to be operationally useful without destroying evidence of origin.

IV. DESIGN PATTERN 1: SINGLE-TABLE MATCHING

Single-table matching is appropriate when the attributes needed to evaluate identity are stored directly on the party record or can be safely rolled up into a denormalized matching view. For example, an individual record may include full name, primary address, phone number, and source identifier in one structure. This pattern reduces configuration complexity and is often effective for compact domains or early proof-of-concept scenarios.

The limitation is that party data is rarely compact at enterprise scale. Customers, suppliers, healthcare professionals, and organizations may have multiple addresses, multiple communication channels, and historical names. Rolling every child attribute into the parent record may simplify matching configuration but can weaken semantic accuracy and make lineage more difficult to explain.

V. DESIGN PATTERN 2: RELATION-BASED MATCHING

Relation-based matching treats child tables as first-class participants in the identity-resolution process. Instead of flattening all names, addresses, and communication channels into the parent party table, the model retains normalized structures and applies match policies to the relevant child entities. The parent party match can then use child-table match results through explicit relationships.

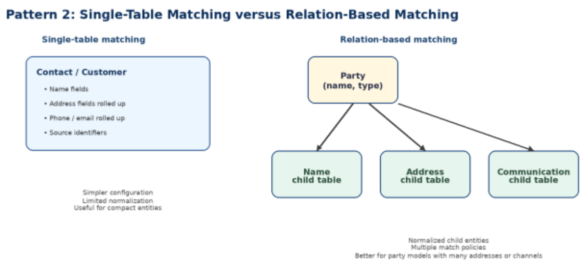

Fig. 2. Comparison of single-table and relation-based party matching.

This pattern is especially useful when records contain multiple addresses or contact channels, or when a party model needs to preserve separate semantics for legal name, preferred name, location, communication, and identifier data. The tradeoff is configuration complexity. Relation-based matching requires careful policy ordering, stable foreign keys, and explicit stewardship views that allow users to inspect the evidence behind a match.

VI. DESIGN PATTERN 3: CONTEXT-AWARE MATCHING

A party model often stores both individual and organization records in a common structure. However, the fields that determine identity are not the same. A person match may rely on first name, last name, full name, date of birth, address, and communication channels. An organization match may rely on legal name, normalized address, registration identifiers, industry classification, and trusted source status. Applying a single universal rule to both entity types creates avoidable risk.

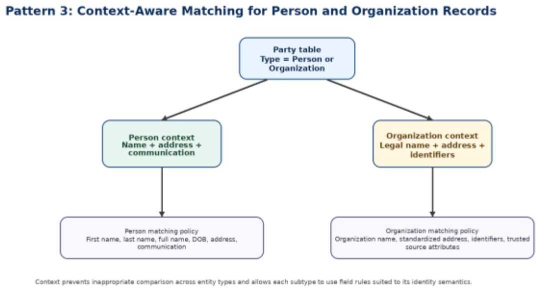

Fig. 3. Context-aware matching separates person and organization identity semantics.

Context-aware matching restricts the boundary of comparison before scoring. A person record is evaluated under person-specific match rules, while an organization record is evaluated under organization-specific match rules. This prevents inappropriate comparisons across entity types and allows each domain to use the evidence that is meaningful for its identity semantics.

VII. DESIGN PATTERN 4: THRESHOLD-GOVERNED STEWARDSHIP

Matching scores should not be interpreted as absolute truth. They are decision-support values that must be interpreted through governance policy. A low score should normally reject a match. A very high score may support automated merge or creation of a golden record candidate. Scores in the middle range should be routed to stewardship because the business risk of false merge or false non-merge is material.

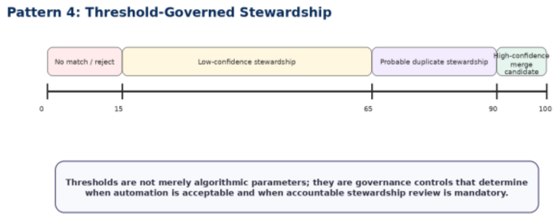

Fig. 4. Threshold-governed stewardship model for suspicious records.

Thresholds should be established jointly by data architects, data stewards, and business owners. They should also be monitored over time. If stewards consistently override automated recommendations, the match policy requires adjustment. If stewards consistently approve the same class of probable duplicates, the organization may be able to increase automation safely.

VIII. SURVIVORSHIP MODEL

Survivorship rules determine how attribute values from source-specific records become the trusted values in the golden record. The best rule is not always most recent or most complete. For some attributes, trusted source hierarchy is more important. For others, validation status, source recency, steward override, or regulatory provenance may dominate. A practical survivorship design should define rules per attribute class rather than applying one global rule.

Pattern 5: Survivorship Decision Matrix

Attribute class	Typical rule	Governance rationale
Identifiers	Preserve source identifiers and map through cross-reference	Traceability and auditability
Legal / party name	Prefer trusted source, then steward override	Business identity accuracy
Address	Standardize, then prefer verified/most recent value	Deliverability and deduplication stability
Communication	Prefer validated phone/email and current status	Operational contact quality
Classification	Use controlled reference data and approval workflow	Consistency across downstream consumers
Audit metadata	Retain all source lineage and match decision history	Explainability and compliance

Fig. 5. Survivorship decision matrix for constructing governed golden records.

The survivorship matrix in Fig. 5 demonstrates the principle that consolidation should retain explainability. The golden record is not merely a new record; it is a governed representation backed by source mappings, match decisions, and survivorship logic. This is critical for audit, troubleshooting, and downstream trust.

IX. IMPLEMENTATION CONSIDERATIONS IN EBX5

A. Policy Ordering

When using relation-based matching, child policies should be configured before parent policies. For example, name, address, and communication policies should be validated before a parent party policy consumes those results. This bottom-up configuration ensures that parent-level identity scoring uses normalized and policy-governed evidence.

B. Stewardship View Design

Stewards should not be forced to interpret opaque scores. The stewardship view should expose candidate records, comparison fields, score components, source systems, and candidate survivor values. The ability to mark records as not suspect, remove records from clusters, add records into clusters, and manually select survivor values is essential for accountable governance.

C. Golden Record Behavior

A consolidation design may either create a new golden record or promote an existing source record as the golden representation. The decision depends on business process, traceability expectations, downstream consumption, and whether source-specific records must remain untouched. In either case, cross-reference preservation is mandatory.

D. False Merge Prevention

False merges are more damaging than missed duplicates in many enterprise contexts because they can corrupt a governed master record and propagate errors to downstream systems. Conservative thresholds, context partitioning, and steward review for ambiguous clusters are practical controls for reducing this risk.

X. EVALUATION CRITERIA

The effectiveness of a matching and survivorship design should be measured across both technical and governance dimensions. Technical metrics include duplicate detection rate, false merge rate, false non-merge rate, average cluster size, and percentage of records automatically resolved. Governance metrics include steward review volume, average decision time, override rate, escalation rate, and number of downstream defects caused by identity resolution errors.

- Duplicate detection rate: percentage of true duplicate groups identified by the policy; used to measure identity-resolution effectiveness.
- False merge rate: percentage of merged records that should not have been merged; used to measure business risk and trust.
- Steward override rate: percentage of steward decisions that reverse an automated recommendation; used as a policy tuning signal.
- Average cluster size: mean number of records presented per suspicious cluster; used to monitor steward workload.
- Lineage completeness: percentage of golden records retaining source cross-references; used to measure auditability.

XI. DISCUSSION

The patterns presented in this paper intentionally combine algorithmic matching with governance controls. This hybrid approach reflects the practical reality of enterprise MDM: automation improves scalability, but stewardship protects semantic correctness. The most successful implementations are not those that maximize auto-merge at all costs, but those that make automation explainable and reserve human judgment for cases where identity evidence remains ambiguous.

The proposed patterns also highlight the importance of domain-specific design. Party consolidation cannot be solved by copying a generic match rule from one domain to another. Person and organization data require different evidence models, thresholds, and survivorship rules. A reusable architecture must therefore support contextual variation while preserving a common governance framework.

XII. CONCLUSION

This paper presented design patterns for matching and survivorship in EBX5 party data consolidation. The framework distinguishes identity resolution from survivorship, compares single-table and relation-based matching, introduces context-aware matching for individual and organization records, and defines threshold-governed stewardship as a control mechanism for ambiguous clusters. The resulting architecture enables organizations to construct governed golden records while preserving lineage, reducing false merges, and making match decisions explainable. These patterns provide a practical foundation for enterprise-scale master data consolidation programs where trust, traceability, and operational repeatability are as important as technical matching accuracy.

XIII. REFERENCES

[1] K. Dutta, “EBX Data Consolidation Solution,” Orchestra Networks technical implementation material, Spring 2018.

[2] K. Dutta, “EBX Standard Operating Procedure (Data Consolidation),” Orchestra Networks, June 2018.

[3] Orchestra Networks, “EBX5 Product and Add-on Documentation,” Orchestra Networks, 2018.

[4] K. Dutta, “Data Exchange Guide to Configure Data Transfer,” Orchestra Networks technical implementation material, 2018.

[5] DAMA International, DAMA-DMBOK: Data Management Body of Knowledge, 2nd ed., Technics Publications, 2017.