

Appendix – The Grumpy Old IT Guy Fuzzy Matching Framework™ (GITG-FMF™)

Version 1.0

Copyright © 2026 Grumpy Old IT Guy Publishers /
Angelo R. Bobak

Executive Summary

The **Grumpy Old IT Guy Fuzzy Matching Framework™ (GITG-FMF™)** is a structured methodology for designing, implementing, and governing fuzzy matching solutions within Microsoft SQL Server environments. The framework provides a repeatable approach for identifying records that have a high percentage of representing the same real-world entity despite differences in spelling, abbreviations, formatting, missing values, typographical errors, or inconsistent data entry practices.

GITG-FMF™ combines traditional SQL Server technologies—including **Data Quality Services (DQS)**, **SQL Server Integration Services (SSIS)**, and **custom Transact-SQL matching routines**—with modern **AI-assisted matching techniques** to improve matching accuracy while maintaining governance, transparency, and human oversight. The framework is intended for customer, supplier, product, employee, patient, citizen, and other

master data domains where duplicate detection is essential to enterprise data quality.

Introduction

Duplicate records remain one of the most common and costly data quality problems facing organizations. Differences in spelling, abbreviations, punctuation, nicknames, transposed characters, incomplete information, and inconsistent formatting often prevent exact matching techniques from identifying records that represent the same business entity.

The **Grumpy Old IT Guy Fuzzy Matching Framework™ (GITG-FMF™)** was introduced in **2026** as part of the Grumpy Old IT Guy family of enterprise data quality frameworks. Its purpose is to provide organizations with a practical methodology for implementing fuzzy matching solutions that are accurate, explainable, scalable, and fully integrated into enterprise data quality and governance programs.

Rather than relying on a single matching algorithm, GITG-FMF™ promotes a layered approach that combines deterministic rules, probabilistic scoring, phonetic comparisons, similarity algorithms, AI-assisted recommendations, and data stewardship review to maximize confidence while minimizing false matches.

Purpose of the Framework

The objectives of GITG-FMF™ are to:

- Identify duplicate business records
- Improve customer and master data quality

- Reduce duplicate creation
- Support Master Data Management (MDM)
- Improve entity resolution
- Increase matching accuracy
- Reduce false positives and false negatives
- Support regulatory and compliance requirements
- Establish repeatable matching governance

Guiding Principles

The framework is built upon seven core principles.

1. Profile Before Matching

Successful fuzzy matching begins with understanding the quality of the underlying data. Data profiling identifies completeness, consistency, formatting issues, and anomalies before matching algorithms are applied.

2. Standardize Before Comparing

Records should be standardized whenever possible. Common abbreviations, casing, punctuation, address formats, telephone numbers, dates, and reference values should be normalized to improve comparison accuracy.

3. Multiple Algorithms Produce Better Results

No single matching algorithm performs well in every situation. GITG-FMF™ encourages combining deterministic, phonetic, statistical, and AI-assisted techniques.

4. Confidence Must Be Measured

Every candidate match should receive a confidence score that reflects the likelihood that two records represent the same entity.

5. Human Stewardship Remains Essential

AI and matching algorithms assist decision-making but do not replace experienced data stewards responsible for approving uncertain matches.

6. Matching Rules Require Governance

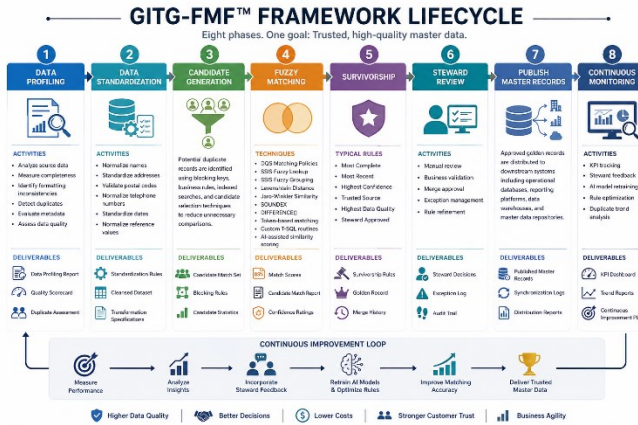
Matching policies should be documented, version controlled, tested, monitored, and periodically reviewed to reflect changing business requirements.

7. Continuous Improvement

Matching quality should improve through ongoing monitoring, steward feedback, and refinement of matching rules and AI models.

Framework Lifecycle

The GITG-FMF™ lifecycle consists of eight phases.



GITG-FMF – Fuzzy Matching Life Cycle

Phase 1 — Data Profiling

Activities include:

- Analyze source data
- Measure completeness
- Identify formatting inconsistencies
- Detect duplicates
- Evaluate metadata
- Assess data quality

Deliverables:

- Data Profiling Report
- Quality Scorecard
- Duplicate Assessment

Phase 2 — Data Standardization

Activities include:

- Normalize names
- Standardize addresses
- Validate postal codes
- Normalize telephone numbers
- Standardize dates
- Normalize reference values

Deliverables:

- Standardization Rules
- Cleansed Dataset
- Transformation Specifications

Phase 3 — Candidate Generation

Potential duplicate records are identified using blocking keys, business rules, indexed searches, and candidate selection techniques to reduce unnecessary comparisons.

Deliverables:

- Candidate Match Set
- Blocking Rules
- Candidate Statistics

Phase 4 — Fuzzy Matching

Candidate records are evaluated using one or more techniques such as:

- DQS Matching Policies

- SSIS Fuzzy Lookup
- SSIS Fuzzy Grouping
- Levenshtein Distance
- Jaro-Winkler Similarity
- SOUNDEX
- DIFFERENCE()
- Token-based matching
- Custom T-SQL routines
- AI-assisted similarity scoring

Deliverables:

- Match Scores
- Candidate Match Report
- Confidence Ratings

Phase 5 — Survivorship

Business rules determine which attributes survive when duplicate records are merged.

Typical rules include:

- Most Complete
- Most Recent
- Highest Confidence
- Trusted Source
- Highest Data Quality
- Steward Approved

Deliverables:

- Survivorship Rules
- Golden Record
- Merge History

Phase 6 — Steward Review

Data stewards review records with uncertain confidence scores.

Activities include:

- Manual review
- Business validation
- Merge approval
- Exception management
- Rule refinement

Deliverables:

- Steward Decisions
- Exception Log
- Audit Trail

Phase 7 — Publish Master Records

Approved golden records are distributed to downstream systems including operational databases, reporting platforms, data warehouses, and master data repositories.

Deliverables:

- Published Master Records
- Synchronization Logs
- Distribution Reports

Phase 8 — Continuous Monitoring

Matching performance is continuously monitored and refined.

Activities include:

- KPI tracking
- Steward feedback
- AI model retraining
- Rule optimization
- Duplicate trend analysis

Deliverables:

- KPI Dashboard
- Trend Reports
- Continuous Improvement Plan

SQL Server Technology Mapping

The framework aligns with Microsoft SQL Server technologies as follows:

Technology	Primary Role
SQL Server Data Quality Services (DQS)	Knowledge-driven cleansing and matching policies
SQL Server Integration Services (SSIS)	ETL workflows, Fuzzy Lookup, and Fuzzy Grouping
Transact-SQL	Custom matching logic, scoring routines, and survivorship processing
SQL Server Agent	Automated matching and scheduling
Master Data Services (MDS)	Golden record management and stewardship

Technology	Primary Role
SQL Server Reporting Services (SSRS) / Power BI	Match quality reporting and dashboards
Microsoft Fabric / Azure	Enterprise integration, analytics, and governance
AI Services	Semantic matching, intelligent recommendations, and continuous learning

Recommended Deliverables

A complete GITG-FMF™ implementation should produce:

- Data Profiling Report
- Matching Requirements Document
- Standardization Rules Catalog
- Candidate Match Register
- Matching Policy Document
- Confidence Score Matrix
- Survivorship Rule Catalog
- Steward Review Log
- Golden Record Register
- Duplicate Resolution Report
- Match Quality Dashboard
- Continuous Improvement Plan

Supporting Templates

The framework includes the following recommended templates:

GITG-FMF-101

Source Data Assessment Checklist

GITG-FMF-102

Data Standardization Rules

GITG-FMF-103

Candidate Match Register

GITG-FMF-104

Matching Rule Specification

GITG-FMF-105

Confidence Scoring Matrix

GITG-FMF-106

Survivorship Decision Matrix

GITG-FMF-107

Steward Review Worksheet

GITG-FMF-108

Golden Record Approval Form

GITG-FMF-109

Duplicate Resolution Log

GITG-FMF-110

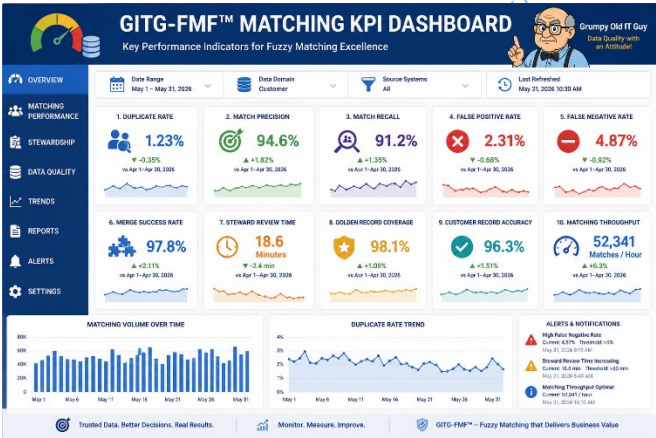
Match Quality KPI Dashboard

GITG-FMF-111

Continuous Improvement Register

Key Performance Indicators

Organizations implementing GITG-FMF™ should monitor:



Matching KPI Dashboard

- Duplicate Rate
- Match Precision
- Match Recall
- False Positive Rate
- False Negative Rate
- Merge Success Rate
- Steward Review Time
- Golden Record Coverage

- Customer Record Accuracy
- Matching Throughput

Relationship to Other GITG Frameworks

GITG-FMF™ complements the broader Grumpy Old IT Guy framework family:

- **GITG-DPF™** provides enterprise data profiling.
- **GITG-DCF™** governs data cleansing and standardization.
- **GITG-RMF™** focuses on enterprise record matching and survivorship.
- **GITG-DQF™** defines the overall data quality methodology.
- **GITG-APF™** extends profiling through AI-assisted discovery and recommendations.
- **GITG-DQMM™** measures organizational data quality maturity.

Together, these frameworks provide a comprehensive methodology for profiling, cleansing, matching, governing, monitoring, and continuously improving enterprise data assets.

Benefits

Organizations adopting GITG-FMF™ can expect:

- Higher duplicate detection accuracy
- Improved master data quality
- Reduced manual matching effort
- Better customer and supplier data
- Increased confidence in analytics
- Stronger governance and auditability

- Lower operational costs associated with duplicate records
- More reliable AI and business intelligence outcomes
- Scalable matching processes across the enterprise

Conclusion

The **Grumpy Old IT Guy Fuzzy Matching Framework™ (GITG-FMF™)** provides a practical, technology-neutral methodology for implementing high-quality fuzzy matching within SQL Server environments while remaining adaptable to modern AI-enabled data platforms. By integrating profiling, standardization, multiple matching algorithms, confidence scoring, survivorship, stewardship, and continuous monitoring, the framework transforms duplicate detection from a technical task into a governed business capability. The result is a sustainable process that produces trusted golden records, strengthens enterprise data quality, and supports more accurate operational, analytical, and AI-driven decision-making.

© 2026 Angelo R. Bobak
All rights reserved.

First Edition: February 2026

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the publisher's prior written permission.

Grumpy Old IT Guy Publishers

Westchester County, NY, USA

www.grumpyolditguy.com

Copyright 2026 - Grumpy Old IT Guy Publishers