

The Grumpy Old IT Guy Customer Data Standardization Framework (GITG-CDSF™)

A Practical Methodology for Standardizing Enterprise Customer Data

Introduction

Customer information is among the most valuable assets an organization possesses. Unfortunately, it is also one of the most inconsistent. Over years of acquisitions, application development, customer self-service, manual data entry, and system migrations, organizations inevitably accumulate multiple versions of the same customer. Variations in names, addresses, phone numbers, company names, abbreviations, spelling errors, and incomplete records reduce the quality of reporting, analytics, customer service, compliance, and operational efficiency.

The **Grumpy Old IT Guy Customer Data Standardization Framework (GITG-CDSF™)** is a practical methodology for standardizing customer information on the Microsoft SQL Server platform. Rather than treating data standardization as a one-time project, the

framework integrates standardization into the organization's continuous data quality lifecycle through repeatable processes, governance, stewardship, automation, and ongoing monitoring.

The framework combines the strengths of:

- SQL Server relational database technology
- SQL Server Integration Services (SSIS)
- Data Quality Services (DQS)
- Master Data Services (MDS)
- Data profiling
- Business rules
- Metadata management
- Data stewardship

Together, these technologies create a governed process that continuously improves customer information throughout the enterprise.

What is Customer Data Standardization?

Customer data standardization is the process of converting customer information into a consistent, validated, and governed format while preserving its business meaning.

Unlike simple data cleansing, standardization establishes enterprise-wide formatting rules that ensure every application represents customer information consistently.

Examples include:

- Converting all state names to approved abbreviations

- Standardizing postal address formats
- Normalizing phone numbers
- Applying consistent capitalization
- Expanding or removing common abbreviations
- Standardizing company suffixes
- Validating postal codes
- Correcting misspellings
- Removing duplicate formatting
- Applying common date formats

The objective is not merely cleaner data, but trustworthy customer information that can be reliably integrated, matched, governed, and analyzed.

Objectives of the GITG-CDSF™

The framework is designed to achieve several strategic objectives:

- Improve customer data consistency
- Reduce duplicate customer records
- Increase matching accuracy
- Improve regulatory compliance
- Simplify enterprise integration
- Improve customer analytics
- Increase operational efficiency
- Establish repeatable governance processes
- Enable trusted master customer records
- Support AI and machine learning initiatives

Core Principles

The framework is based upon seven guiding principles.



Core Principles – Customer Data Standardization

1. Profile Before Standardizing

Never begin standardization without understanding the existing data.

Data profiling identifies:

- Missing values
- Invalid values
- Pattern inconsistencies
- Duplicate records
- Business rule violations
- Outlier values
- Data distributions

Profiling establishes baseline quality metrics that guide subsequent standardization efforts.

2. Standardize Before Matching

Matching algorithms perform significantly better when data has first been standardized.

For example:

Raw Customer Names

IBM Corp.

International Business Machines

I.B.M.

IBM Corporation

After standardization

International Business Machines Corporation

Matching accuracy increases dramatically because formatting inconsistencies have been removed.

3. Business Rules Drive Standardization

The framework relies on business-defined rules rather than ad hoc transformations.

Examples include:

- Approved country codes

- Corporate suffix standards
- Address abbreviations
- Phone number formatting
- Preferred customer naming conventions
- Postal code validation
- State and province abbreviations

Business rules are managed centrally and evolve as business requirements change.

4. Metadata is the Foundation

Every standardization rule should be documented.

Metadata defines:

- Attribute definitions
- Allowable values
- Reference domains
- Validation rules
- Steward ownership
- Transformation logic
- Data lineage

Well-managed metadata ensures consistent implementation across systems.

5. Master Data Governs Standardized Values

After standardization, authoritative customer records should be maintained within a master data repository.

Master records become the enterprise's trusted source for customer information while preserving historical relationships and auditability.

6. Automation Maximizes Consistency

Manual correction cannot scale across millions of records.

Automation ensures consistent execution of:

- SSIS transformation packages
- DQS cleansing projects
- Validation workflows
- Scheduled synchronization
- Exception processing
- Reporting

7. Continuous Improvement

Customer data continuously changes.

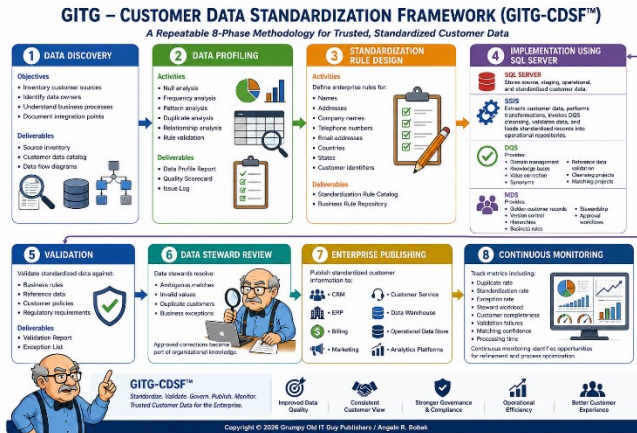
The framework incorporates ongoing:

- Monitoring
- Profiling
- Rule refinement
- Steward review
- Exception management
- Performance measurement

Customer standardization becomes an operational capability rather than a one-time project.

GITG-CDSF™ Methodology

The framework consists of eight repeatable phases.



Customer Data Standardization Framework Methodology

Phase 1 – Data Discovery

Objectives

- Inventory customer sources
- Identify data owners
- Understand business processes
- Document integration points

Deliverables

- Source inventory
- Customer data catalog
- Data flow diagrams

Phase 2 – Data Profiling

Activities

- Null analysis
- Frequency analysis
- Pattern analysis
- Duplicate analysis
- Relationship analysis
- Rule validation

Deliverables

- Data Profile Report
- Quality Scorecard
- Issue Log

Phase 3 – Standardization Rule Design

Activities

Define enterprise rules for:

- Names
- Addresses
- Company names
- Telephone numbers
- Email addresses
- Countries
- States
- Customer identifiers

Deliverables

- Standardization Rule Catalog
- Business Rule Repository

Phase 4 – Implementation Using SQL Server

Technology Responsibilities

SQL Server

Stores source, staging, operational, and standardized customer data.

SSIS

Extracts customer data, performs transformations, invokes DQS cleansing, validates data, and loads standardized records into operational repositories.

DQS

Provides:

- Domain management
- Knowledge bases
- Value correction
- Synonyms
- Reference data validation
- Cleansing projects
- Matching projects

MDS

Provides:

- Golden customer records
- Version control
- Hierarchies
- Business rules

- Stewardship
- Approval workflows

Phase 5 – Validation

Validate standardized data against:

- Business rules
- Reference data
- Customer policies
- Regulatory requirements

Deliverables

- Validation Report
- Exception List

Phase 6 – Data Steward Review

Data stewards resolve:

- Ambiguous matches
- Invalid values
- Duplicate customers
- Business exceptions

Approved corrections become part of organizational knowledge.

Phase 7 – Enterprise Publishing

Publish standardized customer information to:

- CRM
- ERP

- Billing
- Marketing
- Customer Service
- Data Warehouse
- Operational Data Store
- Analytics Platforms

Phase 8 – Continuous Monitoring

Track metrics including:

- Duplicate rate
- Standardization rate
- Exception rate
- Steward workload
- Customer completeness
- Validation failures
- Matching confidence
- Processing time

Continuous monitoring identifies opportunities for refinement and process optimization.

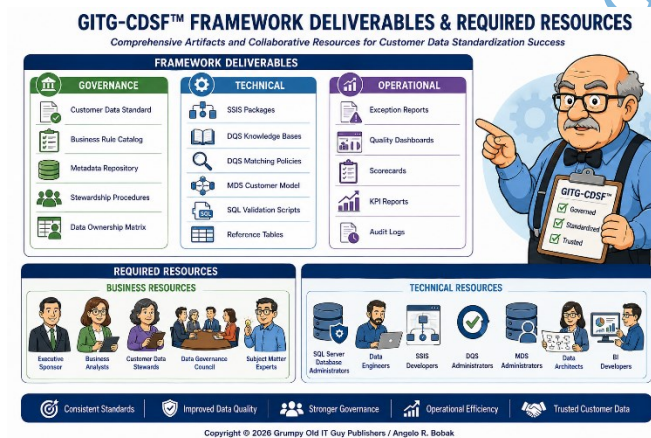
The **Grumpy Old IT Guy Customer Data Standardization Framework (GITG-CDSF™)** provides a structured, repeatable methodology for transforming inconsistent customer information into standardized, governed, and trusted enterprise data.

Through eight integrated phases—Data Discovery, Data Profiling, Standardization Rule Design, SQL Server Implementation, Validation, Data Steward Review, Enterprise Publishing, and Continuous Monitoring- the framework combines business rules, governance, stewardship, and SQL Server technologies to improve

customer data quality. The methodology establishes standardization as a continuous operational capability that supports reliable matching, integration, analytics, and enterprise decision-making.

Framework Deliverables

Implementation of the GITG-CDSF™ produces a comprehensive set of governance and operational artifacts.



Customer Data Standardization Framework Deliverables

Governance

- Customer Data Standard
- Business Rule Catalog
- Metadata Repository
- Stewardship Procedures
- Data Ownership Matrix

Technical

- SSIS Packages
- DQS Knowledge Bases
- DQS Matching Policies
- MDS Customer Model
- SQL Validation Scripts
- Reference Tables

Operational

- Exception Reports
- Quality Dashboards
- Scorecards
- KPI Reports
- Audit Logs

Required Resources

Successful implementation requires collaboration across business and technology teams.

Business Resources

- Executive Sponsor
- Business Analysts
- Customer Data Stewards
- Data Governance Council
- Subject Matter Experts

Technical Resources

- SQL Server Database Administrators
- Data Engineers
- SSIS Developers
- DQS Administrators
- MDS Administrators

- Data Architects
- BI Developers

Benefits of the Framework

Organizations implementing the GITG-CDSF™ can expect improvements in several areas.

Operational Benefits

- Higher customer data consistency
- Fewer duplicate records
- Improved customer search accuracy
- Faster onboarding
- Reduced manual corrections
- Simplified integrations

Business Benefits

- Better customer analytics
- Improved marketing segmentation
- More accurate reporting
- Enhanced customer experience
- Improved regulatory compliance

Governance Benefits

- Defined ownership
- Repeatable processes
- Measurable quality improvements
- Auditable changes
- Sustainable stewardship

Relationship with the GITG Data Quality Framework

The GITG-CDSF™ is one component of the broader **Grumpy Old IT Guy Data Quality Framework (GITG-DQF™)**.

Within the overall data quality lifecycle:

- **Data Profiling** identifies customer data issues.
- **Customer Data Standardization (GITG-CDSF™)** applies enterprise formatting, validation, and normalization rules.
- **Data Cleansing (GITG-DCF™)** corrects data defects and resolves quality issues.
- **Record Matching and Survivorship** identify duplicate customer entities.
- **Master Data Management** establishes authoritative customer records.
- **Data Governance** ensures ongoing ownership, stewardship, policy enforcement, and continuous improvement.

Together, these frameworks provide a comprehensive methodology for transforming inconsistent customer information into trusted, governed, and reusable enterprise data assets.

Conclusion

Customer data standardization is the foundation upon which effective customer master data management, analytics, artificial intelligence, and enterprise reporting are built. The Grumpy Old IT Guy Customer Data

Standardization Framework (GITG-CDSF™) provides a structured, repeatable, and technology-enabled methodology for achieving consistent customer information across the enterprise. By integrating SQL Server, SSIS, DQS, and MDS with governance, stewardship, metadata, and business rules, organizations can establish a sustainable customer data standardization capability that continually improves data quality, reduces operational risk, and enables trusted decision-making across every business function.

© 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