

The Grumpy Old IT Guy Data Cleansing Framework (GITG-DCF™)

A Structured Framework for Enterprise Data Cleansing

Version 1.0

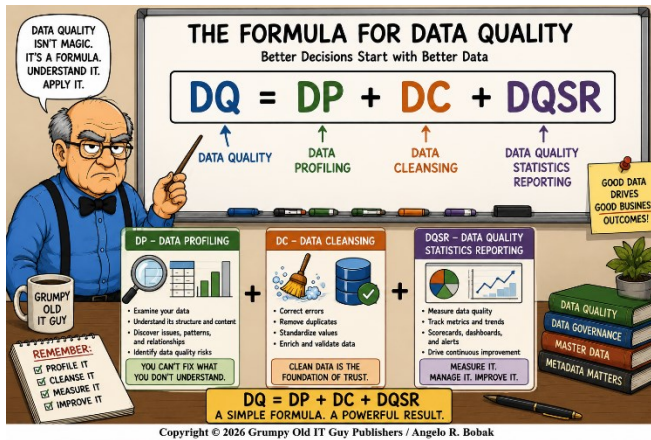
**Copyright © 2026 Grumpy Old IT Guy Publishers /
Angelo R. Bobak**

Executive Summary

Data has become one of the most valuable assets within modern organizations. Yet despite significant investments in data warehouses, analytics platforms, artificial intelligence, and cloud technologies, many organizations continue to rely on inaccurate, inconsistent, duplicated, and incomplete data.

Poor data quality directly affects operational efficiency, customer satisfaction, regulatory compliance, and executive decision-making.

This framework applies to all critical business data: customers, products, addresses, manufacturers, and other key business dimensions.



The **Grumpy Old IT Guy Data Cleansing Framework (GITG-DCF™)** provides a practical and repeatable methodology for systematically identifying, correcting, standardizing, validating, and monitoring enterprise data.

The Grumpy Old IT Guy Data Quality Formula

The **Grumpy Old IT Guy Data Quality Formula** represents data quality as a continuous progression from understanding data to improving and measuring it:

DQ (Data Quality) = DP (Data Profiling) + DC (Data Cleansing) + DQSR (Data Quality Statistics Reporting)

Data Profiling identifies the conditions, patterns, anomalies, and weaknesses within the data. **Data Cleansing** corrects and standardizes the problems discovered during profiling, while **Data Quality Statistics Reporting** measures the results through metrics, trends, scorecards, and KPIs.

Together, these three disciplines create a continuous improvement cycle: **profile the data, cleanse what is wrong, measure the results, and repeat.** In the Grumpy Old IT Guy approach, trusted data is not created by a single cleanup effort—it is maintained through continuous assessment, correction, measurement, and improvement.

Unlike traditional cleansing projects that treat data correction as a one-time activity, the GITG-DCF™ recognizes that data quality is a continuous operational discipline. Cleansing must become an integral component of enterprise data governance, stewardship, metadata management, master data management, and continuous monitoring.

The framework provides organizations with a structured roadmap that transforms raw, inconsistent data into trusted enterprise information assets.

What is the GITG Data Cleansing Framework?

The **Grumpy Old IT Guy Data Cleansing Framework (GITG-DCF™)** is a structured methodology that defines the processes, governance, standards, technologies, and operational activities required to improve and maintain enterprise data quality.

The framework performs cleansing at three distinct levels:

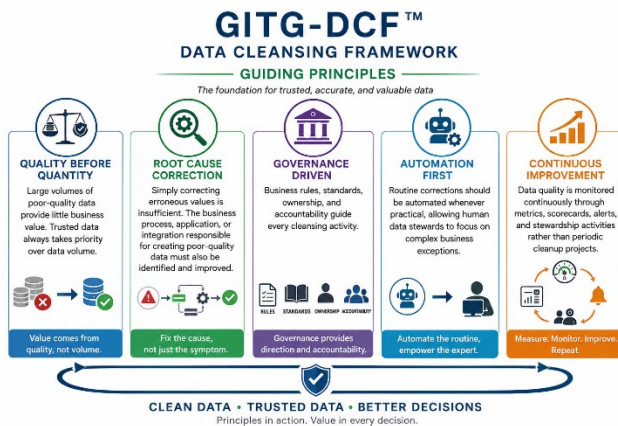
- **Attribute Level** – correcting individual data elements
- **Record Level** – correcting entire business records

- **Dataset Level** – improving quality across complete databases, warehouses, data lakes, or analytical platforms

Rather than viewing cleansing as a single project, GITG-DCF™ embeds cleansing into the organization's overall Data Quality Lifecycle where profiling, governance, stewardship, remediation, monitoring, and continuous improvement operate together.

Guiding Principles

The framework is built upon several core principles.



Data Cleaning Framework Guiding Principles

Quality Before Quantity

Large volumes of poor-quality data provide little business value. Trusted data always takes priority over data volume.

Root Cause Correction

Simply correcting erroneous values is insufficient. The business process, application, or integration responsible for creating poor-quality data must also be identified and improved.

Governance Driven

Business rules, standards, ownership, and accountability guide every cleansing activity.

Automation First

Routine corrections should be automated whenever practical, allowing human data stewards to focus on complex business exceptions.

Continuous Improvement

Data quality is monitored continuously through metrics, scorecards, alerts, and stewardship activities rather than periodic cleanup projects.

Framework Objectives

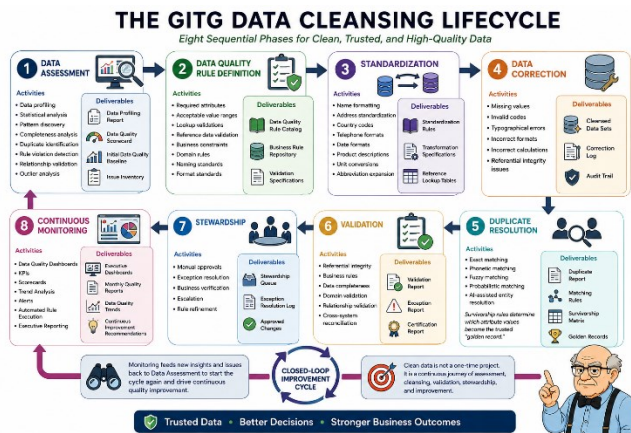
The GITG-DCF™ seeks to:

- Improve data accuracy
- Eliminate duplicate records
- Standardize business values
- Increase completeness
- Improve consistency across systems
- Validate business rules
- Strengthen regulatory compliance
- Increase confidence in reporting
- Improve AI and analytics readiness

- Reduce operational costs caused by bad data quality

The GITG Data Cleansing Lifecycle

The framework consists of eight sequential phases.



Data Cleansing Lifecycle

Phase 1 – Data Assessment

Every cleansing initiative begins with understanding the current condition of enterprise data.

Activities include:

- Data profiling
- Statistical analysis
- Pattern discovery
- Completeness analysis
- Duplicate identification

- Rule violation detection
- Relationship validation
- Outlier analysis

Deliverables

- Data Profiling Report
- Data Quality Scorecard
- Initial Data Quality Baseline
- Issue Inventory

Phase 2 – Data Quality Rule Definition

Business and technical stakeholders define the rules that determine acceptable data quality.

Examples include:

- Required attributes
- Acceptable value ranges
- Lookup validations
- Reference data validation
- Business constraints
- Domain rules
- Naming standards
- Format standards

Deliverables

- Data Quality Rule Catalog
- Business Rule Repository
- Validation Specifications

Phase 3 – Standardization

Data values are converted into consistent enterprise formats.

Examples include:

- Name formatting
- Address standardization
- Country codes
- Telephone formats
- Date formats
- Product descriptions
- Unit conversions
- Abbreviation expansion

Deliverables

- Standardization Rules
- Transformation Specifications
- Reference Lookup Tables

Phase 4 – Data Correction

Incorrect values are corrected using predefined business rules, reference data, or authoritative master data sources.

Typical corrections include:

- Missing values
- Invalid codes
- Typographical errors
- Incorrect formats
- Incorrect calculations
- Referential integrity issues

Deliverables

- Cleansed Data Sets
- Correction Log
- Audit Trail

Phase 5 – Duplicate Resolution

Duplicate business entities are identified and consolidated.

Matching techniques include:

- Exact matching
- Phonetic matching
- Fuzzy matching
- Probabilistic matching
- AI-assisted entity resolution

Survivorship rules determine which attribute values become the trusted "golden record."

Deliverables

- Duplicate Report
- Matching Rules
- Survivorship Matrix
- Golden Records

Phase 6 – Validation

After cleansing, all corrected data is revalidated against enterprise quality standards.

Validation includes:

- Referential integrity
- Business rules

- Data completeness
- Domain validation
- Relationship validation
- Cross-system reconciliation

Deliverables

- Validation Report
- Exception Report
- Certification Report

Phase 7 – Stewardship

Business Data Stewards review exceptions that cannot be automatically corrected.

Typical stewardship activities include:

- Manual approvals
- Exception resolution
- Business verification
- Escalation
- Rule refinement

Deliverables

- Stewardship Queue
- Exception Resolution Log
- Approved Changes

Phase 8 – Continuous Monitoring

The framework concludes with ongoing operational monitoring.

Monitoring includes:

- Data Quality Dashboards
- KPIs
- Scorecards
- Trend Analysis
- Alerts
- Automated Rule Execution
- Executive Reporting

The monitoring phase continuously feeds new observations back into the assessment phase, creating a closed-loop improvement cycle.

Deliverables

- Executive Dashboards
- Monthly Quality Reports
- Data Quality Trends
- Continuous Improvement Recommendations

Framework Deliverables

Implementation of GITG-DCFTM produces a comprehensive collection of reusable assets, including:

- Enterprise Data Cleansing Strategy
- Data Quality Assessment Report
- Data Profiling Results
- Data Quality Rule Catalog
- Cleansing Standards
- Standardization Dictionary
- Reference Data Repository
- Master Data Mapping
- Matching Rules

- Survivorship Rules
- Cleansing Procedures
- Exception Management Procedures
- Stewardship Workflows
- Data Quality KPIs
- Executive Dashboards
- Monitoring Reports
- Continuous Improvement Roadmap

Required Resources

Successful implementation requires collaboration across business and technical teams.

Executive Sponsor

Provides organizational support, funding, and strategic direction.

Data Governance Council

Approves enterprise policies, standards, and priorities.

Data Owners

Define business meaning and approve quality standards.

Data Stewards

Review exceptions, approve corrections, and maintain ongoing quality.

Data Quality Analysts

Perform profiling, analysis, and rule development.

Database Administrators

Implement cleansing processes and optimize performance.

Data Engineers

Develop ETL pipelines, transformations, automation, and integrations.

Application Owners

Resolve source system issues and implement preventive controls.

Enterprise Architects

Ensure alignment with enterprise information architecture.

Technology Components

Although the framework is technology-independent, common supporting platforms include:

- SQL Server
 - Data Quality Services (DQS)
 - Master Data Services (MDS)
 - SQL Server Integration Services (SSIS)
 - Microsoft Purview
 - Fabric Data Factory
 - Power BI
 - SQL Reporting Services
 - Data Catalogs

- Metadata Repositories
- AI-assisted profiling tools
- Machine Learning models
- Enterprise ETL platforms

Expected Benefits

Organizations implementing GITG-DCF™ typically experience measurable improvements in:

- Data accuracy
- Customer record quality
- Regulatory compliance
- Reporting consistency
- Operational efficiency
- Analytics reliability
- AI model performance
- Reduced duplicate records
- Reduced manual correction effort
- Improved executive confidence in enterprise data

Relationship with the Grumpy Old IT Guy Data Quality Framework

The GITG-DCF™ is one component of the broader Grumpy Old IT Guy Data Quality ecosystem.

The relationship can be summarized as:

**Data Profiling → Data Cleansing → Data Quality
Measurement → Governance → Monitoring →
Continuous Improvement**

Each framework builds upon the previous one, creating an integrated enterprise data quality architecture that continuously measures, improves, and governs information assets.

Conclusion

High-quality data is not achieved through occasional cleanup efforts. It is the result of disciplined governance, standardized processes, accountable stewardship, intelligent automation, and continuous monitoring.

The **Grumpy Old IT Guy Data Cleansing Framework (GITG-DCF™)** provides organizations with a practical methodology for transforming inconsistent, incomplete, and duplicated data into trusted enterprise information. By integrating cleansing into the broader data quality lifecycle, organizations establish a sustainable foundation for analytics, regulatory compliance, artificial intelligence, and informed decision-making.

Rather than asking, "How do we clean our data?" the framework encourages organizations to ask, "How do we continuously maintain trusted data?" That shift in perspective transforms data cleansing from a reactive task into a strategic organizational capability.

© 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