

Quality Gating: Where Trustworthy Master Data Begins

IBM Master Data Management (MDM) is a powerful matching engine. Its results are only as strong as the data fed into it. **At scale, undetected data quality issues lead to overlays, orphaned records, and super-entities.**

IBM Knowledge Catalog (IKC) acts as a quality gate, assessing data before it enters MDM to help **enforce quality thresholds upfront and reduce remediation from months to hours.**



Quality gating is not governance theater. It is risk containment.

How MDM Benefits From Additional Data Quality Control

IBM Master Data Management (IBM MDM) creates a single, trusted view of your most critical data, such as patients, providers, employees, or locations. By ingesting records from multiple source systems, comparing attributes, and linking them into a unified golden record, **MDM eliminates data silos and underpins confident, enterprise-wide decision-making.**

INTELLIGENT RECORD LINKING



- ▶ Accurately **connects records across disparate systems** using **sophisticated probabilistic matching**, even with limited overlap.

THE GOLDEN RECORD



- ▶ Produces a **single, authoritative view** of every entity across the enterprise, acting as the **system of record for critical data.**

ENTERPRISE SCALE



- ▶ Designed to handle **tens of millions of records** with **high-performance matching and real-time search.**

MDM's matching algorithms operate on records **exactly as they arrive**. This deliberate design delivers exceptional accuracy when records are reliable. This means the quality of what you load directly shapes the quality of the golden records MDM produces.

IBM MDM historically succeeds in linking records across imperfect source systems, but at true enterprise scale, **proactive measurement and profiling of data becomes essential to protecting your data investment.**

Three Data Quality Risks at Scale



When data quality risks persist at enterprise scale, their effects quickly compound and spread across the organization.

Consequences at Scale

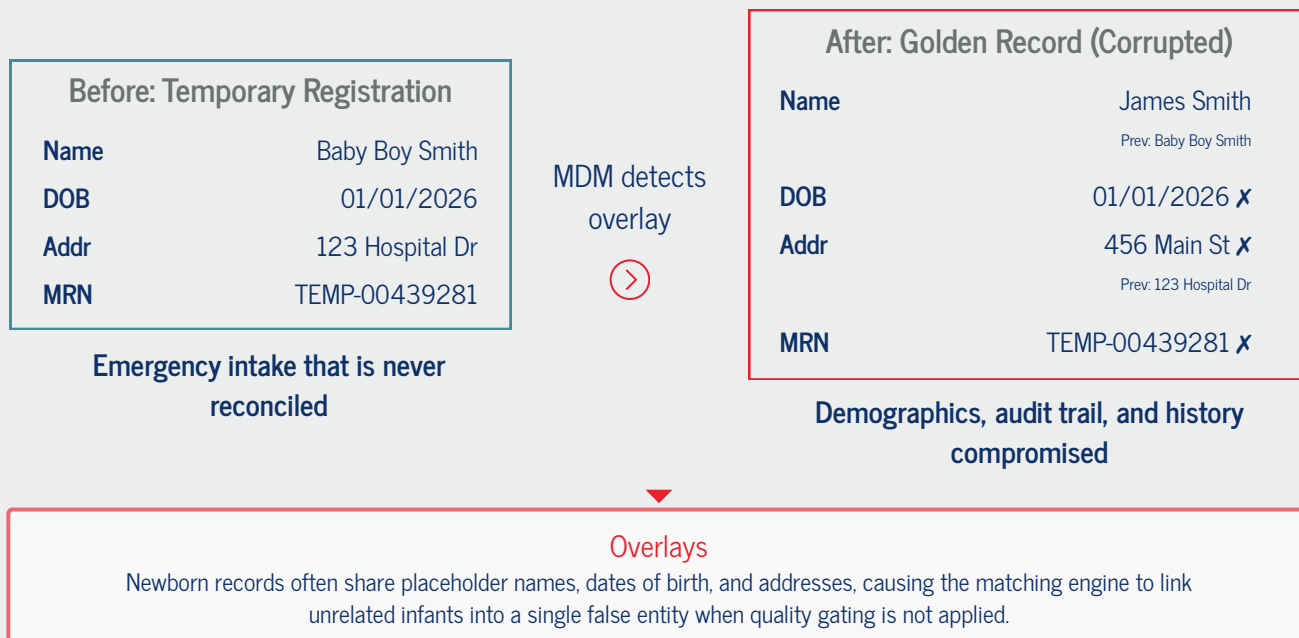
- Trust in master data erodes as identities become unreliable
- Errors propagate downstream, impacting billing, reporting, analytics, and operations
- Remediation consumes months of effort, often requiring large-scale reprocessing
- Costs rise sharply as manual data steward intervention increases
- Business value is delayed while teams focus on cleanup instead of outcomes

At massive volumes of records, even a **small percentage of problematic data** can affect **hundreds of thousands of entities**. Three patterns account for the vast majority of post-load remediation work.

SCENARIO 1: OVERLAYS

WHEN ONE PERSON'S DATA OVERWRITES ANOTHER'S

An overlay occurs when one individual's record is inadvertently merged with a completely different person's data. Unlike a duplicate, an overlay is a data overwrite. The original information is replaced rather than just copied, making it **difficult to detect and costly to reverse**.



With **quality gating**: placeholder values are detected and blocked before they reach MDM.

SCENARIO 2: SUPER-LARGE ENTITIES

WHEN PLACEHOLDERS BECOME MATCH SIGNALS

A super-large entity (sometimes called a mega-entity or spider entity) is a golden record that has grown to encompass thousands of unrelated records. MDM provides anonymous value configuration specifically to prevent this. Once a placeholder is registered, the engine excludes it from matching logic.

The risk emerges when the anonymous value configuration step is incomplete before data is loaded, and the matching engine treats those values as legitimate demographic signals.

Example with anonymous values across different attributes:

SSN: 999-99-9999
SSN: 9999
SSN: 999-99-9998

Addr: PO Box 1
Addr: Unknown
Addr: 1 Govt Plaza

DOB: 01/01/1900
DOB: 01/01/1800

+ hundreds of thousands more records
share these values

Matching engine links records sharing field values

Super-Large Entity

Hundreds of thousands of unrelated individuals merged into one golden record
Untangling a single super-entity can take extensive time and cost a vast amount of resources

With **quality gating**: placeholder values are identified as anonymous and excluded from matching before they can collapse unrelated records into a super-entity.

SCENARIO 3: ORPHANED RECORDS

DATA THAT CAN NEVER MATCH

When records arrive with extremely sparse or low-quality key attributes, they cannot be reliably matched to any real-world entity and remain permanently orphaned. Keeping these unmatchable records in MDM silently **drives up storage, compute, and licensing costs while degrading performance**.

Example 1: Emergency Intake PII

Name	Unknown Unknown
DOB	01/01/1900
Addr	Unknown, CA 90000
MRN	P-999999

Patient with no valid PII

Example 2: Newborn Placeholder

Name	Baby Girl [Placeholder]
DOB	04/29/2026 [Today]
Mother's ID	[Not Reconciled]
MRN	[Null]

Temporary record never updated

Example 3: Placeholder Lab Results

Name	Baby Boy
Phone	04/29/2026 [Today]
Lab ID	L-001
Email	dummy@email.com

Record with dummy contacts

Orphaned Records

These unmatchable records silently accumulate, degrading search and match performance while inflating storage, compute, and licensing costs.

With **quality gating**: systematically invalid attributes are detected early, by enforcing minimum data quality and completeness thresholds, preventing orphaned records from entering MDM and accumulating indefinitely.

IBM Knowledge Catalog as Your Quality Gate

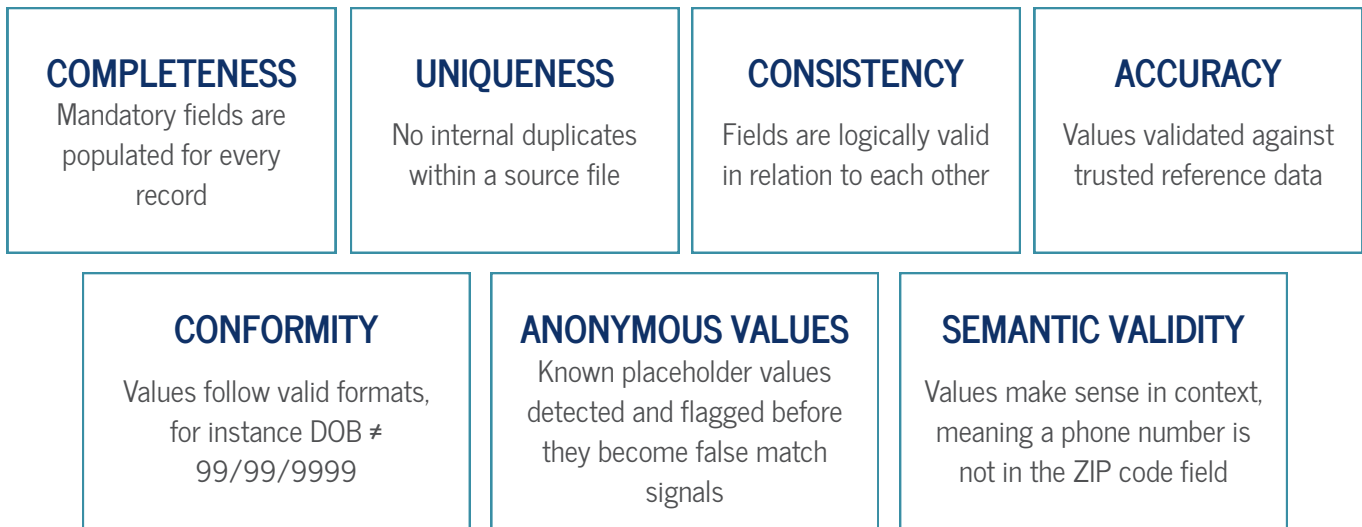
IBM Knowledge Catalog (IKC) is IBM's data governance and quality management platform. When introduced into the MDM ingestion pipeline, IKC can evaluate each incoming file against defined quality rules, and holds data back until it meets your standards. **No record will reach MDM without passing the gate.**

The IKC Quality Gate Pipeline



SEVEN QUALITY DIMENSIONS

IKC scores incoming files across the dimensions below, assessing an overall quality percentage for each file. Files that fall below the defined quality threshold, are held for review before any data enters MDM.

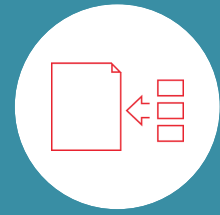


When a file is held, organizations have **three options**:

- Block the entire file for offline remediation
- Selectively load only the records that pass
- Route failing records through automated correction workflows using IBM QualityStage

Only qualifying records proceed to MDM ingestion.

USE CASE



HEALTHCARE MIGRATION: MILLIONS OF RECORDS ACROSS MULTIPLE SOURCE SYSTEMS

A large healthcare organization migrating patient records from **multiple systems** establishes a **quality threshold** of 87%. Before a single record enters MDM, the quality of source file is assessed. The results will reveal **stark data quality differences** and **prevent two problems** that would be much more costly to fix after the fact.

Source System	Quality Score	Status
EMR System A	93.2%	✓ Approved
EMR System B	88.7%	✓ Approved
Legacy Hospital Registration (System C)	71.4%	✗ Blocked
Emergency Dept. Temporary Registration Feed	58.9%	✗ Blocked

ISSUE 1: LEGACY REGISTRATION SYSTEM (71.4%)

IKC identifies hundreds of thousands of records with placeholder names, and tens of thousands of records with default dates of birth, both high-risk factors for super-entity formation. The load is blocked, and data owners are notified immediately.

ISSUE 2: EMERGENCY DEPARTMENT FEED (58.9%)

Blocking the entire ED feed was not operationally viable, as active emergency registrations needed to remain accessible. IKC's selective loading may be applied here.

OUTCOME: WITH THE QUALITY GATE

- ▶ Source records corrected and anonymous value rules validated before load
- ▶ Super-entity formation prevented entirely
- ▶ Extended post-load remediation timelines avoided

OUTCOME: WITH SELECTIVE LOADING

- ▶ High-quality records load MDM immediately
- ▶ Partial records route to a reconciliation queue for data steward review
- ▶ Placeholder-only records are staged rather than discarded, pending source system updates
- ▶ Emergency data remains operationally available throughout the migration

The Cost of Deferring Quality Assessment

Across industries, organizations that load data without prior quality validation face similar remediation timelines. **These patterns are predictable, and almost entirely preventable.**

- **Thousands** of customer accounts overlaid into a single dataset during a financial services migration
- **Months** of remediation required to resolve the impact of the overlay event
- **Millions** of loyalty records orphaned due to a single shared invalid email address
- **Vast numbers** of individuals merged into a single "golden record," creating super-entities
- **Severe** compliance risks emerge, including potential HIPAA violations from mislinked PHI
- **Extensive** manual remediation needed to correct multiple complex super-entity records

These are not platform failures.

They are predictable consequences of loading unvalidated data at scale in any MDM implementation.

Quality assessment upstream converts multi-month remediation projects into hours-long correction cycles.

Extending the Value of Your IBM MDM Investment

IBM MDM is designed to be the **authoritative system of record** for your organization's most critical data. The organizations that realize its full potential are those that treat **upstream data governance as an enabler, not an obstacle.**

THREE PRACTICES THAT ENABLE TRUSTED GOLDEN RECORDS:

ANONYMOUS VALUE CONFIGURATION

Register all known placeholder values so the matching engine excludes them as demographic signals to prevent super-entity formation at the source.

SOURCE-SPECIFIC QUALITY THRESHOLDS

Set thresholds appropriate to each source system's data profile and criticality. Different systems warrant different standards.

DEFINED REMEDIATION WORKFLOWS

Establish clear paths for held records, such as automated correction via QualityStage, selective loading, or source-system feedback loops so quality issues are **resolved instead of just blocked.**

Organizations that invest in these practices before the first data load typically discover and resolve issues in hours to days. **Those that defer quality assessment often spend months to over a year on remediation after the fact.**

IMT Can Help With Your Data Quality Strategy

Whether you are preparing for an initial MDM data load, assessing an existing environment, or working through data quality issues, **IMT brings deep experience across MDM implementation, data governance, and IBM Knowledge Catalog configuration.**

MDM IMPLEMENTATIONS



End-to-end IBM MDM deployments with quality governance built in from day one.

DATA QUALITY GOVERNANCE



IKC rule design, quality threshold modeling, and remediation workflow setup.

ENVIRONMENT REMEDIATION



Pragmatic strategies for untangling super-entities, resolving overlays, and recovering orphaned records.

The selective loading approach described above allows high-quality records to load immediately, while questionable data is set aside for later reconciliation. **This is a practical solution for organizations that cannot afford to reject entire data sources.**

With more than 30 years of experience in data management across industries, systems, and data domains, IMT helps organizations apply this approach in ways that fit their real-world constraints and business goals.

Contact IMT to discuss your MDM data quality strategy.
imt.ca/connect