

**Pathway to master data
management with master data
integration from IBM.**

Contents

- 2 Part 1: What is MDM and why is it important?**
- 6 Part 2: Functional components of master data management**
- 16 Part 3: Pathway to master data management**
- 25 Part 4: Best practices, benefits & success**

In our complicated world of enterprise business applications running on complex information technology architectures, we often come full circle to discover “hot” topics that were the same things we started out to do years ago. Each original set of core applications supporting basic front and back office functions has always managed its own reference data about core business entities. In today’s information intensive environment, the reference data about an enterprise’s customers, vendors, items and products must be accurate, shared across functions, and available “on demand” for a variety of heterogeneous applications. “Master Data Management” with its acronym label “MDM” is receiving a lot of attention as an “enterprise” business and technology opportunity to manage shared reference data that deserves serious consideration.

Master data management is really a hybrid software solution. In one sense the idea of MDM is already included in the legacy and packaged enterprise application portfolio. On the other hand it’s fundamentally a data integration capability that needs to be available in real time across applications manipulating the facts about your core business entities. Whether it’s reducing cost, meeting legal requirements for compliance, exchanging “standard” data with partners, implementing new applications or consolidating old ones, master data needs to be processed in a consistent fashion. Best practices suggest methodology to drive data integration development and related business process alignment is critical to success. Framed by data integration and business process methodology the overall MDM solution needs to leverage the current legacy and enterprise applications footprint. Master data integration services bind master data management applications with a variety of business applications and provide the integration services required to acquire, transform and deliver master data in real time throughout the enterprise.

This white paper takes a close look at the information integration necessary for successful MDM. First we'll try to describe exactly what MDM means and why it's emerging so strongly as a CXO topic. Then we'll describe the functional components of a complete master data management solution with a focus on requirements for master data integration. Next up is a look at alternatives for deployment and operational architectures. In closing we'll highlight the critical role of methodology, discuss benefits and review customer success with MDM solutions from IBM.

Part 1: What is it and why is it important?

To start let's suggest a working definition: Master data management is the business process, applications and technical integration architecture used to create and maintain accurate and consistent views of core business entities across disparate applications in the enterprise. In other words, how a financial institution ensures that no matter how many ways you interact with them, e.g., ATM, internet, branch office, credit card, marketing and custom service departments, they know you. But isn't that what we've been doing all along? Do you have one view into your customers, suppliers, employees, products and materials that reflects the facts out there in the real world – without data duplication and inaccuracies?

Unfortunately problems related to the enterprise management of “master” data are more acute than ever before. Creation and transmission of data describing core business entities such as customers, products and locations continues to grow exponentially within and outside of all organizations. Most large businesses lack a consistent accurate “system of record” for master data. In reality, many application silos hold fragmented versions of master data, but no one application has the complete picture. At the same time data quality continues to emerge as a legitimate business issue driven by inter-operability and compliance requirements. Adding to the complexity, a new generation of browser based portal style applications drive requirements for more discrete “master” data sets to be delivered in real time.

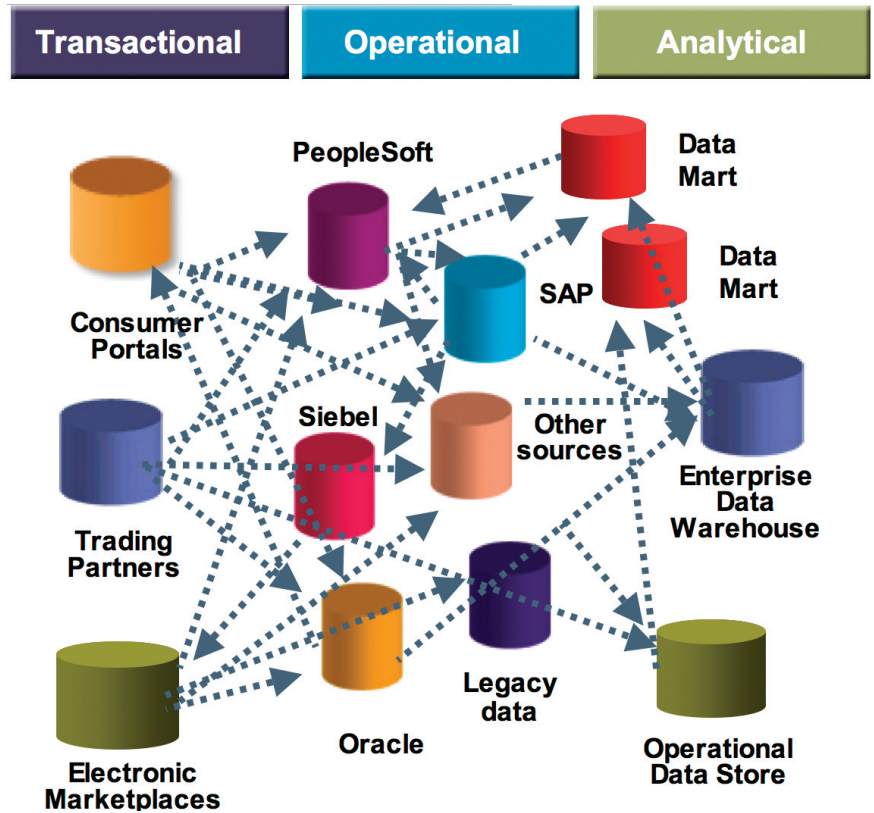


Figure 1. Enterprise Complexity Managing Master Data

For example, in the high-technology manufacturing industry, it can take up to four weeks to get a new product added to various provisioning systems, local territory applications and partner systems before one new unit can be sold and shipped. Many of these companies are investing in MDM to help them introduce new procedures and processes to reduce new product introduction to a matter of days.

Translation: at the very time your business requires instant access to high quality information about your core business entities, the facts that describe them are scattered to the far corners of more data transaction structures, databases, datamarts and spreadsheets than you can even count – and locations in your enterprise you didn't know existed!

No wonder you have eleven million customer identification numbers but only eight million customers. The lack of a single consistent enterprise view of core business entities severely impacts the enterprise.

Here's the evidence:

“CRM less than what we were planning for”

Why? There was never a logically correct view of who the customers are, how they're related, where they're located and what they've bought from us.

- **Profitability:** *66% indicate profitability of company as a whole was negatively affected by poor information quality.¹*
- **Customer service:** *75% indicate bad customer data quality is harming customer service, quality and loyalty.*
- **Inaccurate:** *52% identified integration of diverse systems as a major source of inaccurate information.*
- **Incorrect:** *A telecommunications firm lost \$8M a month because data entry errors incorrectly coded accounts, preventing bills from being sent out.*

“Items can't synchronize”

Why? We purchase, manufacture and inventory in 100 locations and there's no common key to link the same material, part or product across the dozens of disparate data sources short of looking at the description of each individual item. And you need a business expert to do it!

- **Errors in data** – *30% of data in retailers systems is wrong*
- **Lost productivity** – *25 minutes manual cleansing per SKU per year*
- **Slow time to market** – *4 weeks to introduce new products*
- **Invoice deductions** – *43% of invoices result in deductions*
- **Failed scans** – *up to 70,000 per week (1 large US Retailer)*
- **Lost sales** – *up to 3.5% per year*

“Business intelligence wrong answers”

Why? Unique keys assigned to primary dimensions like customer, vendor and product do not reflect the facts out in the real world. When you aggregate the basics like units and dollars, the answers are incorrect and do not agree with answers generated from other sources.

- **Limited acceptance** – *Through 2007 more than 50% of data warehouse projects will suffer limited acceptance, if not outright failure, as a result of lack of attention to data quality issues*
- **Productivity loss** - *In most BI efforts approximately 80% of the IT effort will be expended on dealing with issues of infrastructure; data analysis, data acquisition and data quality*

There are many more. Vendors we cannot consolidate. Members, subscribers and providers we can't cross reference. Loan exposure we can't accurately describe. Regulations and standards we can't comply with.

So back to the buzz around MDM. Master data management is the opportunity to:

- *Implement an information integration platform that can access the facts about your core business entities from anywhere in your enterprise*
- *Automate the creation of a single logically correct view based on your business rules that agrees with the facts in the real world*
- *Actively manage master data repositories with master data management systems*
- *Automate and standardize business processes such as “introduce new product” or “set up new customer”*
- *Deliver high quality master data to your current suite of business applications in real time*

Sure we've been doing this all along. But the opportunity to leverage your master data in your next generation of IT architecture and drive concrete benefits for a wide range of critical business initiatives is worth better understanding now. In Part 2 we'll dissect all the functional components that make up the Master Data Management Solution. In Part 3 we'll look at architectural alternatives to deployment and operational integration with the current application portfolio. Part 4 focuses on the critical role of methodology, discusses benefits and closes with examples of customer success in MDM with IBM.

Part 2: Functional components of master data management

Master data management is really a hybrid offering. On one hand, MDM is used by business users such as brand managers to manage their portfolio of product items and financial planners to manage their charts of account. On the other hand MDM is fundamentally part of the integration fabric that enables consistent access to and management of core business entities such as customers, suppliers and vendors across the enterprise. At the heart of a successful master data management solution are the capabilities to automate the determination if one "instance" of a master data entity is the same as another. For example, in order to process a purchase, when presented with a new "instance" of a customer (comprised of the basic facts that describe the customer like name and address) the MDM solution must decide, based on your business rules, if this customer already exists or is new to the enterprise. Sounds easy. Not exactly!

IBM Master Data Management

Gain Control over Comprehensive Master Information and Business Processes

- Speed deployment and ROI with master data solutions, industry and content models
- Manage comprehensive master information and business processes
- Leverage industry leading master data integration capabilities
- Deep expertise spanning business transformation, best practices and methodologies gained from 100's of successful deployments
- Strong partner ecosystem

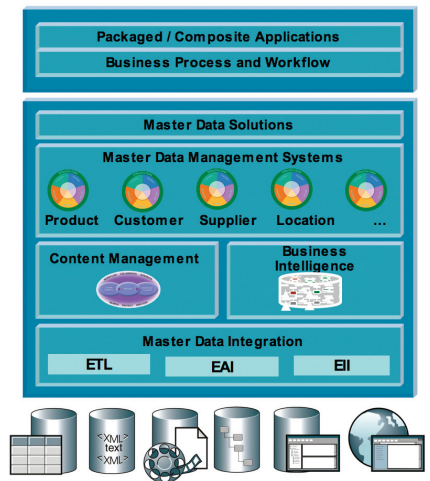


Figure 2. IBM Master Data Management

The IBM approach to providing a MDM solution includes the following major components:

I. Master data management system – a transactional software infrastructure that manages a repository of reference data through a package of business services. These services includes administrative and user screens, repository management functionality, workflows and events that define and maintain the data hierarchies, relationships and attributes associated with specific data elements and industries. For example:

- *For Retailers and Consumer Packaged Goods manufacturers, an MDM system such as IBM WebSphere® Product Center manages part numbers, descriptions, pricing, images and packaging details, suppliers and vendors. This form of MDM system is commonly referred to as Product Information Management (PIM).*

- *For Financial Services, Retailing Banking, Telecommunications, and Insurance companies, an MDM system such as IBM WebSphere Customer Center, manages financial profiles, location, demographic data, billing and account information. This type of MDM system implements a concept of “party” that covers all roles associated with individuals and legal entities and is commonly referred to as Customer Data Integration (CDI).*
- *Additionally, organizations often build custom applications for MDM to provide product or customer oriented information management.*

2. Master data integration - provides a single integration infrastructure necessary to solve implementation and ongoing operational data challenges across any industry, business function and scope of business data. This includes managing master data business transactions and ensuring the data is synchronized across the enterprise. In addition, federated middleware enables MDM systems to dynamically access external data sources for content such as images and documents related to the entities managed by the MDM system.

3. Master data solutions – beyond a horizontal technology, MDM comprises industry specific data models, workflows and business processes that ensure the speedy customization of MDM systems targeted at core business domains. Additionally, this includes reusable integration templates that help to rapidly move and cleanse data more quickly from common application sources such as ERP and CRM during the initial load of the MDM repository.

Master data root cause analysis

Before a master data repository can be deployed it first needs to be created with a single logically correct view of the entity, such as customer or product, from across the enterprise. Unfortunately there is no agreement on the meaning or “semantics” of how the facts about customer are stored in applications across the enterprise. Establishing an agreed upon meaning for the facts that describe a master data entity is critical to the successful establishment and continued integration of any master data repository.

Root Cause Analysis: Inconsistent Master Data Across Enterprise Sources

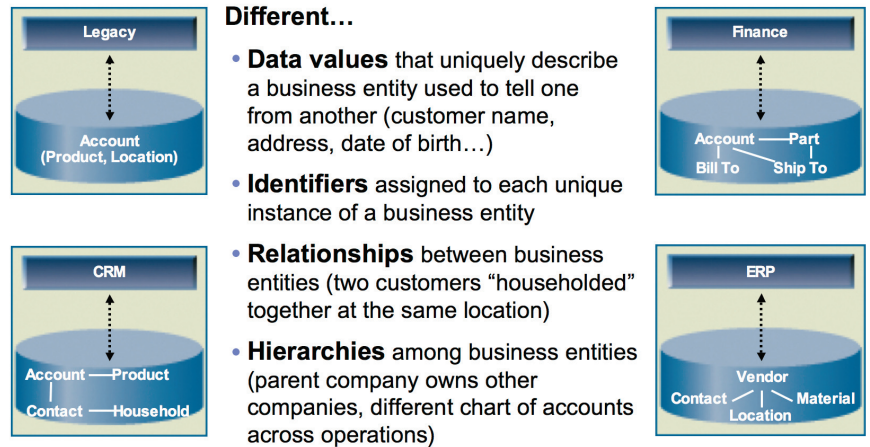


Figure 3. Root cause analysis

For example, the attributes used to describe customers throughout the enterprise usually do not have a common definition. The piece parts of name (first, last, title, salutation ...) and location (street number, street name, street type, suite, apartment, PO Box, city, state, zip...) may have no stricter definition than “name line one 40 characters long.” The same is true of attributes related to product (name, brand, size, color, package, weight...) dumped into “product description 50 characters long.”

Each application that maintains data about core business entities like customer, location, vendor, part, product and material does it in terms of the purpose of the application (using its own data model) - but this is never consistent across the enterprise’s systems. “Legacy” applications are often account oriented with no particular notion of customer master data entities often described as “name and address lines 1 through 5.” Finance applications care if the “Bill To” customer is credit worthy and will pay for goods delivered and services received. ERP applications are all about who to contact to ship a specific part, product or material to a particular location. CRM applications care about customer contacts, how customers are related to each other and how to optimize the contacts across touch points.

Projects that require master data by definition must draw data from across the various application silos. All too often organizations fail to pay proper attention to understanding source data before it's transformed and moved. Poor quality data cause projects to fail. Don't make the mistake of assuming these data quality myths are true.

- *We know our data*
- *Source documentation exists and is accurate*
- *Our metadata accurately describes the data*
- *Our users use the systems "appropriately"*
- *The business rules have not changed over time*
- *Subject matter experts are available*
- *Project scope provides adequate data quality assessment & re-mediation time*

Successful master data management logically binds the master data across disparate applications by creating a repository of core reference data within the Master Data Management system. Cross-reference keys link the objects in the MDM system and their equivalents in each of the other applications, databases or external reference sources. It's these cross-reference keys that enable the enterprise to work with one version of the truth that reflects the facts out in the real world.

This very basic process of "re-engineering" the data facts, uncovering relationships between core business entities, creating and maintaining logically correct keys, managing those keys via cross-references and leveraging a common data integration platform is critical to success.

Master data integration (MDI) Overview

Let's look at the functional components necessary to transform facts from anywhere in your enterprise into high quality master data at initial deployment and for on-going data integrity of the MDM system.

Master Data Management Functional Stack

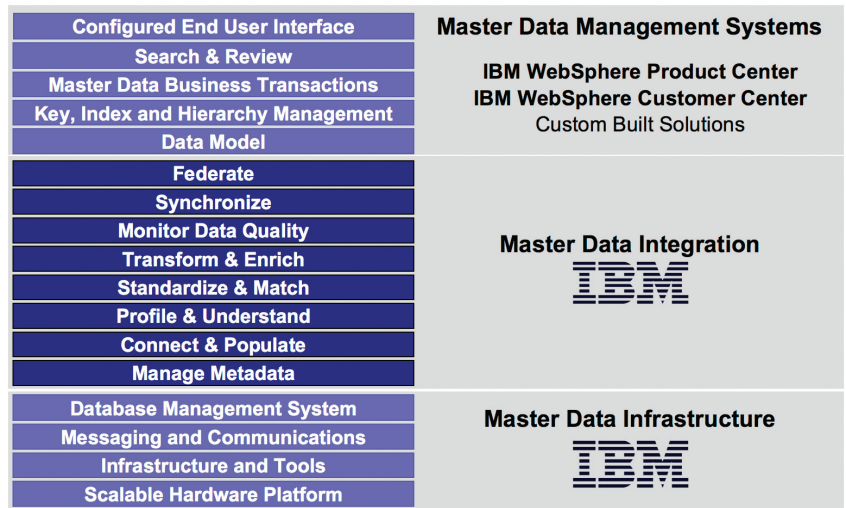


Figure 4: Master data management functional stack

Profile & understand

First of all, it is vital to get an understanding of the data within source systems. This helps establish the quality of the data as it exists today, the usage patterns and rules associated with the data. Data profiling tools automate this process and provide terrific insights into source system data without the laborious manual efforts of traditional data analysis. IBM WebSphere ProfileStage™ is an example of an automated profiling solution that provides column, cross-column and cross-table analysis. It also allows data analysts to define a metadata map of source systems that can be used by downstream processes to determine and assemble linked records. The metadata can be used as the basis for a new master data model, if one is being derived from source system schemas, or simply to map source schemas to a new target model.

Once the content and structure of source systems is understood, the data needs to be put into a working area where heavy interactions will not impact source systems. Typically, this involves the creation of a temporary staging database. This database provides a copy of the data in the source systems that can be used to build cleansing, matching,

and cross-reference creation logic. Please refer to the IBM white paper *Profiling: Take the first step to assuring data quality*² for detailed coverage of the whys, hows and benefits of source data profiling.

Standardization

With the staging database in place, rules can be created for standardizing the facts related to core business entities found on records sourced from throughout the enterprise. Standardization is the process of “fix fielding” the facts or attributes and assigning a business oriented semantic label to each fix fielded fact. Standardization for master data must be flexible to accommodate processing the variety of individual name, commercial name, international location / language, product, part and material data on a common platform. Standardization ensures that the facts are understood in a business context and lays the groundwork for matching and record-linkage.

Global address verification & certification

With today’s global economy many organizations seek to consolidate their understanding of core business entities from around the world. This means that worldwide location information must be standardized in a consistent fashion using local business rules. Then when the piece parts of location have been standardized it’s important to know if those location facts represent a legitimate address. Global address verification matches the standardized facts about each instance of location against reference databases compiled by third parties to verify that a street location in a particular city and state is valid. Certification software confirms an additional postal discount offered by postal authorities for customers who use software they have “certified” as performing high quality address standardization such as WebSphere QualityStage™ CASS software certified by the US Postal Service.

Matching & record linkage

Matching and record linkage is a challenging technical problem that takes one set of facts (bob rich 17 briar tuck road) related to an instance of a master data entity (customer) and compares those facts to a reference store to determine if that instance (robert rich 17 briarhill

rd) of the entity is already known. Matching takes place because there is no common key available across the enterprise. The matching process must see through data anomalies and conflicts. In practical terms, take one set of input data that describes a customer and in ½ a second, with 99.5% accuracy, automatically determine whether that customer exists on the customer master file. Traditionally this work was so difficult, resource and time intensive, that organizations could only do it offline periodically and in many cases farmed the whole problem out to a service bureau. What's driving master data management today as an enterprise business strategy is the need to match and link records about core business entities in real time across a variety of enterprise applications.

From an architecture perspective many organizations look at this as a common “service” that implements a consistent set of matching and linkage rules in batch, near real time and real time for a variety of applications and business purposes. A deep dive exploring matching technology is available through the IBM white paper *IBM WebSphere QualityStage: Superior technology produces superior results.*³

Master data enrichment

Enriching master data is the process of linking standardized facts around core business entities such as individual and commercial customer to third party reference data to extend and expand the understanding of each instance of the customer. For example, many organizations extend their view of individual customers with demographics and psychographics from vendors like Acxiom with InfoBase, just as commercial attributes like revenue, SIC Code and legal relationships between establishments are available from vendors like Dun & Bradstreet. Leveraging third party data as a core competency of the master data management function adds significant value to applications that use the data.

Survivorship

The output from the matching and record linkage process is often expressed as a set; a group of records, each with facts that the matching technology determined describes the same entity in the real world. The assignment

of a single key to that group creates the “logical key integrity” that in turn provides the foundation for a “360 view” and a “single version of the truth”. Survivorship is the business of deciding which combination of facts is “best” for a particular business purpose. As we’ll discuss a little further down, sometimes all that is maintained is the cross reference from the legacy keys in the set to the “group” key assigned to the set. Most often the records from the set are combined to create a better view of that customer or product or part. The individual facts can be selected based on the values that occur most frequently, are most recent or come from a particular source. Just like standardization and matching rules, survivorship rules are flexible for different business purposes and can be applied a transaction or batch at a time on the output from the match process.

Transformation

MDM often involves extracting large amounts of data out of source systems and transformation to a common format, in both batch and per-transaction scenarios. Ideally, these rules should be defined once and applied consistently across both scenarios. IBM WebSphere DataStage® allows rules to be defined for transformation and delivery that can be used in both batch and real-time scenarios across multiple projects. For example, WebSphere DataStage supports the intensive processing requirements of moving very large bulk data sets and the complex transformation requirements for creating analytical views of data for load to a data warehouse and other enterprise targets, such as an MDM system.

Synchronize

Synchronization is the process of maintaining consistency across the master data store and all source systems. Synchronization is a very complex process, since it involves keeping track of changes to any source and to the master data store. When a change occurs, it needs to be replicated to all affected systems, according to survivorship rules (which are typically governed by the data quality solution). During the synchronization process, things like transactional integrity are very important. If one system gets updated, it is important that all systems receive that update as quickly as possible, to avoid discrepancies, and if one update fails, the others

should be rolled back. This level of process synchronization requires specialized technologies such as IBM WebSphere DataStage TX and IBM WebSphere Business Integrator that are capable of integrating with any type of source system and can ensure transactional integrity. The synchronization process is very much tied into the message bus and process event architecture of the business. Consequently, a close examination of where to splice the MDM systems' processes supported by the MDI layer into the existing flow of data between applications needs to occur. Effectively, the MDM system becomes another Publisher and Subscriber of data.

Federate

In many situations, the deployment of a MDM system motivates multiple follow-on projects. With an established reliable source of reference data the MDM deployment can be used as a basis to serve a host of new applications. For example, in the case of products, an organization might decide to extend the value of the MDM solution by supporting an e-commerce site and the creation of a print catalog. It's easy to see why having a centralized repository of product information would be the logical starting point.

However in both scenarios, these solutions require a range of other on-demand information, such as product photos, real-time availability information, affinity data like "customers also purchased these items." Access to this information outside the MDM system can be coordinated and presented simultaneously via federation. Federation provides a mechanism to query multiple sources of information simultaneously returning related information across applications, mainframe databases such as VSAM, IMS™ and DB2® Universal Database™, and content sources such as IBM DB2 Content Manager, FileNet and Documentum. For example, pulling Robert Rich's contract history up from the data warehouse, querying for call details made to the support center and retrieving his photo from the HR system about contractors.

Taken together these are the functional capabilities for data integration necessary to support a complete master data management solution. We see integration of data around master data as a core

competency organizations must develop in support of a variety of enterprise initiatives. We refer to these “master data integration services” as the data integration layer that lives on top of a base infrastructure in support of a “master data management systems.”

Part 3: Pathway to master data management

Master data management is truly an enterprise level initiative that integrates horizontally across the enterprise. Many of the most significant challenges with successful MDM are not technical, but related to process and organization. In fact, MDM is truly a journey undertaken by the enterprise to integrate process, applications and underlying technology in pursuit of core objectives related to efficiency, effectiveness, return on investment and competitive differentiation. The “pathway” to MDM needs to incorporate methodology and consider architectural alternatives for deployment and operation.

Critical role of data integration methodology

A critical success factor for all data integration projects, especially MDM, is methodology. Methodology organizes and binds:

- **Process** – *the steps (and relationships between steps) necessary to build and implement data integration solutions*
- **Technology** – *the functional components applied at each process step to manipulate data*
- **Resources** – *the skill sets required to apply the technology and perform the process step with clearly defined roles and responsibilities*
- **Deliverables** – *the artifacts produced by the process steps and phases that taken together automate the acquisition, transformation and delivery of data*

IBM MDI Methodology – Critical for Successful MDM

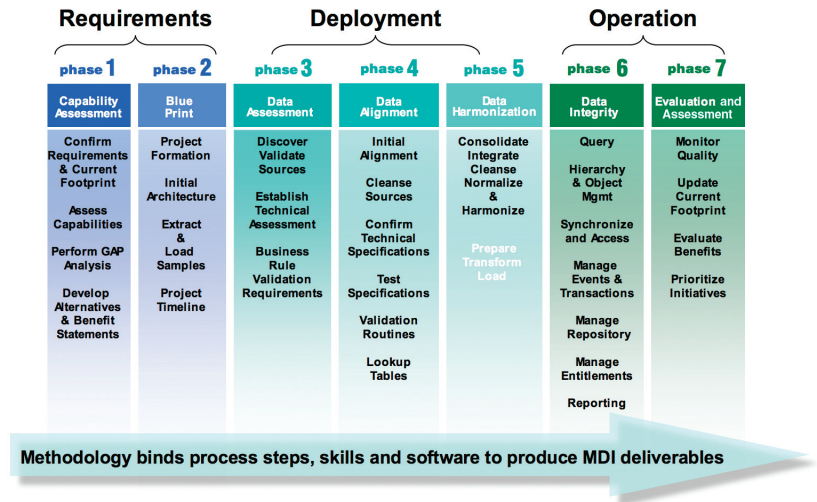


Figure 5. IBM master data integration methodology

Methodology makes successful master data integration a reality! IBM’s Information Integration Solutions Iterations® Methodology is a proven framework for data integration infrastructure projects, based on the knowledge gained from more than 500 projects by IBM consultants and partners. It’s a comprehensive, step-by-step roadmap that takes your team from initial planning and strategy, all the way through tactical implementation and production deployment.

IBM’s Iterations® Methodology includes the following high level process steps for design and implementation of master data integration services:

- **Capability assessment & blueprint** ensures that projects are grounded in the reality of the current state of the business / IT infrastructure, reflect an understanding of the most significant opportunities and are formulated in the context of a roadmap with an overarching vision.
- **Data assessment & validation** understands on a source by source basis the current state of master data from a technical and business rule perspective.
- **Data alignment** defines the processing necessary to acquire, standardize, validate and map master data on a source specific basis for initial loads, batch changes, transactions and real time updates.

- **Data harmonization** defines the deduplication, enrichment from reference sources, record linkage, consolidation, survivorship, transformation, aggregation and mapping specifications for master data across sources for initial loads, batch changes, transactions and real time updates.
- **Data integrity, evaluation and assessment** specifies the process, measurements, maintenance mechanisms and reporting for the master data in an operational setting and closes the project loop by examining benefits achieved and lessons learned.

Methodology is also critical from a business process perspective. It's often the case opportunities to optimize the business process need to be leveraged in parallel with the data integration effort in order for the complete solution to be effective in production.

All of the functional capabilities related to master data integration are consumed by the methodology as pictured in Figure 6.

IBM MDI Functions by Phase

Capability Category	Requirements		Deployment			Operation	
	phase 1	phase 2	phase 3	phase 4	phase 5	phase 6	phase 7
	Capability Assessment	Blue Print	Data Assessment	Data Alignment	Data Harmonization	Data Integrity	Evaluation and Assessment
Federate			✓		✓	✓	
Synchronize						✓	
Monitor Data Quality						✓	✓
Transform & Enrich				✓	✓	✓	
Match & Standardize				✓	✓	✓	
Profile & Understand	✓	✓	✓	✓	✓	✓	
Connect & Populate		✓	✓	✓	✓	✓	
Manage Metadata				✓	✓	✓	

Figure 6. Master data integration capabilities by phase

IBM offers the broadest range of information integration products to implement and operate master data management solutions.

IBM WebSphere Products for Master Data Integration

Capability Category	Requirements		Deployment			Operation	
	phase 1	phase 2	phase 3	phase 4	phase 5	phase 6	phase 7
	Capability Assessment	Blue Print	Data Assessment	Data Alignment	Data Harmonization	Data Integrity	Evaluation and Assessment
Federate	WebSphere Information Integrator Standard and Content Editions						
Synchronize	WebSphere Business Integrator, DataStage TX, II Replication Edition						
Monitor Data Quality			WebSphere AuditStage				
Transform & Enrich			WebSphere DataStage and QualityStage				
Match & Standardize			WebSphere QualityStage				
Profile & Understand			WebSphere ProfileStage				
Connect & Populate			WebSphere DataStage				
Manage Metadata			WebSphere MetaStage				

Figure 7. IBM Products for master data integration

A complete master data management solution has many different components, several primary logical designs and a variety of operational architectures. In all cases, given the nature of master data, it's important to set up a data integration infrastructure using a data integration methodology.

Master data management “styles”

Active management of master data goes beyond sending extracts to a service bureau or even maintaining master data “offline” in a warehouse. A complete MDM solution is integrated operationally and provides a logically correct view of master data to applications and business processes in real time. There are three basic deployment styles:

- *Cross Reference*
- *Light Gold Copy*
- *Extended Gold Copy*

Cross-reference

The cross-reference approach provides a lightweight internal repository for authoritative master data information. Gartner refers to this as the “Registry Style” of customer data integration². Cross-references leverage data quality technology to uniquely identify matching customer records within and across systems. The cross-reference database assigns a primary key to each unique record and stores the identifiers of all linked records in all source systems. It also stores enough identifying data for the customer, location or product to allow matching identification when incomplete records are received and primary keys are not known. As updates are made to source systems, matching services ensure that duplicates are not created. When an inbound record is identified as a match, existing data on the customer can be assembled together from source systems on the fly, based on the linkage information and the survivorship rules. In addition, any updates that effect the cross-reference database can be captured and enforced in near real-time.

Creating this cross-reference is one of the most challenging aspects of MDM. It requires developing a strong understanding of source systems and implementing business driven matching and survivorship rules. The engine used to load the data must be capable of working through large amounts of data in all the source systems. It provides a detailed record of how records within different source systems are related, without losing the overriding the context of any of the sources. It also must be capable of maintaining a metadata lineage of the sources and processing of the data.

“Light” Gold Copy

“Gold Copy” means the recognized “official” version for master data such as customer. The “light” version of the gold copy covers all of the attributes necessary to differentiate one instance of an entity like customer from another. If the business decides that there are 40 elements that uniquely define an individual customer (all the piece parts of name and location plus phone, email, SSN, etc.) then the gold copy becomes the “system of record” for all those elements. Otherwise this is exactly the same scope as the Cross Reference.

Extended Gold Copy

In this style the MDM schema covers everything the business knows about the entity including all of the cross reference and light gold copy attributes and relationships. With customer this is often referred to as Customer Data Integration (CDI) and can include the complete purchase, promotional, service and payment history of the customer as well as customer preferences and staged offers that leverage the insight generated by business intelligence applications. IBM WebSphere Customer Center is an example of an extended gold copy of any “party” such as customer, employee and organization. This is sometimes referred to as the “Transaction Hub” and is the most comprehensive implementation of CDI.

This approach is conceptually similar to the cross-reference approach, but instead of just storing the cross-reference, it stores a complete unique record for each customer. This approach creates a copy of the best data from different source systems and stores that copy in an operational data store that can then be used as an authoritative source of customer information. It works similarly to the cross-reference approach, utilizing data quality technology to uniquely identify, match, and link records within and across systems. In addition, it optionally stores a single consolidated record based on survivorship rules from across the various sources. In order to be operationally effective, this master data store must be kept synchronized with source systems. This requires that processes are put in place to ensure this synchronization. These processes are triggered by updates in any individual source system, and optionally from updates to the master database.

The extended gold copy approach assumes a complete customer data model that is appropriate to represent all business requirements. This data model may be derived from source system schemas, representing a “superset” of those schemas, or it may be an industry or solution-specific model. For example, if you are implementing IBM’s WebSphere Customer Center or WebSphere Product Center solution, the model is delivered as part of the solution and sources are simply mapped to this model. It is likely that this model will still benefit from additional analytics provided by a data warehouse, so there may still be an external linkage to this data.

Extended gold copy is the most complete solution to the master customer data problem. The creation of a master customer information database will deliver the most accurate, up-to-date and complete single view of the customer across multiple channels and business lines in heterogeneous IT environments for the foreseeable future. However, a complete implementation of this approach can be very challenging for many organizations. Building the synchronization and management processes can be one of the most challenging aspects of this approach.

When using a packaged solution from a vendor, it is still important to rationalize a set of master customer data across systems prior to implementation. Packaged solutions emphasize the need for master data rationalization because their function is the syndication and management of master customer data objects.

MDM deployment framework

Regardless of the master data entity and operational style, deployment covers the same four steps. This is also the same process and data integration components used for a variety of enterprise initiatives.

MDM Deployment

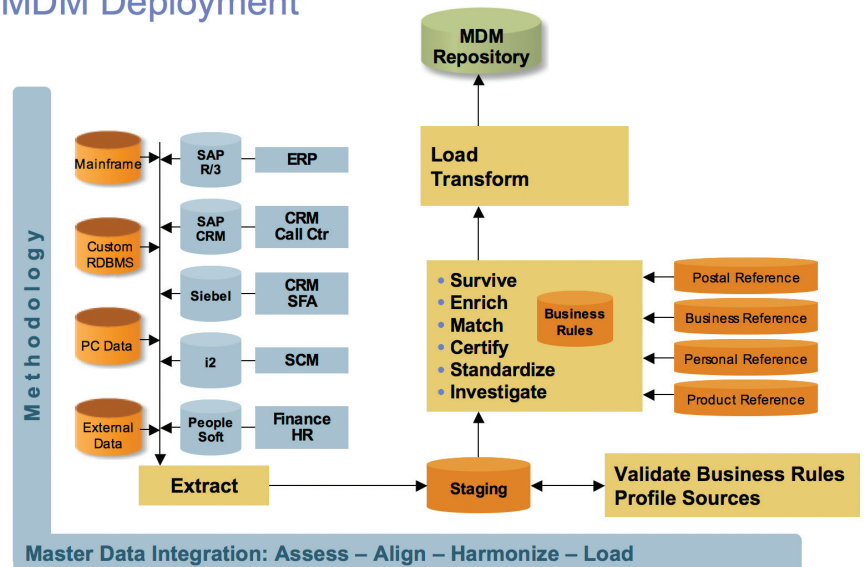


Figure 8. MDM deployment

1. Extract data into staging area

First data must be extracted into a “safe harbor” for assessment and subsequent transformation.

2. Profile & align individual sources

As discussed earlier, it’s critical to first establish an understanding of source data via profiling in order to confirm target and transformation requirements, as well as uncover business practices that only surface upon close inspection of the data values themselves. Alignment provides a vehicle to re-mediate each source system one at a time. Often there’s an opportunity to look at changes to source system applications, process flow and responsibilities to improve the quality of the application’s data.

3. Standardize, verify, match, enrich and survive

Now the data can be transformed, literally re-engineered starting with the data values to produce the high quality facts, logical keys, relationship and hierarchies required to populate a master data repository. During project development the business confirms the rule sets that produce the structure, content and quality of data required by the master data solution. These business driven business rules form the foundation to maintain high levels of data quality in production environments.

4. Transform & load MDM repository

Finally high quality data is moved into the master data repository. The data is transformed and formatted in terms of the particular requirements of the target. Often times data is loaded directly into the structures managed by the repository. Other times the output from the transformation process will be in an XML format so that the business services layer of the master data management solution can process the load as input transactions and write directly to the underlying schema.

MDM operational framework

Making master data and the related management of the information operational for all three deployment styles requires two additional components; inline cleansing and synchronization.

MDM Operational Framework

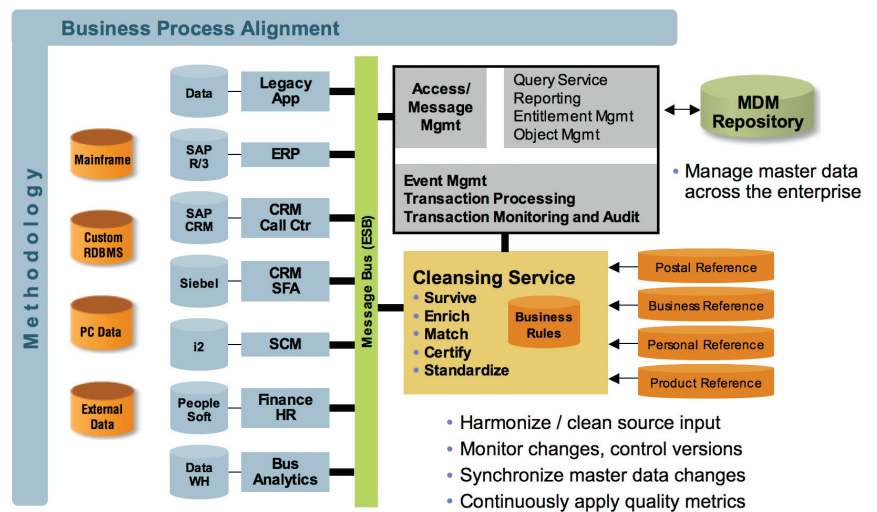


Figure 9. MDM operational framework

Cleansing & matching services

Matching services ensure the ongoing integrity of data. These services take the rules defined in the data quality technology and make them available to be called on the fly by any application that creates master data. This ensures that new data entering the system is not producing a duplicate record and that it is appropriately linked to existing records. Loosely coupled service oriented architecture provides the additional benefit of making the same cleansing services available directly to the master data management system or optionally to other applications independent of the MDM system. For example, the party matching logic for CDI can be shared by marketing applications. These services optionally allow new information within an inbound duplicate record to be “survived” into a master record. For example, new address data may be received for an existing customer. We don’t want to create a duplicate record, but we also don’t want to lose the new address.

Synchronization services

Synchronization is the process of maintaining consistency across the master data repository and all source systems. Synchronization is a very complex process, since it involves keeping track of changes to any source

and to the master data repository. When a change occurs, it needs to be replicated to all affected systems, according to survivorship rules (which are typically governed by the data quality solution). During the synchronization process, transactional integrity is very important. If one system gets updated, it is important that all systems receive that update as quickly as possible, to avoid discrepancies, and if one update fails, the others should be rolled back. This level of process synchronization requires specialized technologies that are capable of integrating with any type of source system and can ensure transactional integrity.

Operationally, new source systems and batch updates are also continually applied to the master data repository in a fashion similar to the deployment model. The transformation and load model often works in reverse so that MDM reference data can be propagated to other solutions such as business intelligence and fraud detection systems.

Part 4: Best practices, benefits & success

Best practices

IBM recommends best practices that can help ensure that your master data management project is a success.

- **Build a detailed business case** *and ensure you have completed a value analysis prior to going into the project. The considerable investment of time and resources required for these projects and the number of departments and groups required to participate make this mandatory.*
- **Align the project initiatives with business objectives** *and ensure executive sponsorship of those objectives. Make sure that your projects support the corporate strategy and try to encourage executives to provide incentives to align the various involved business units and functions.*
- **Design the project to produce quick hits with measurable ROI.** *Use prototyping and proofs of concept where possible to avoid long product cycles and maintain business involvement throughout the process.*

- **Understand the “master data” interaction lifecycle.** *Examining the full people / process lifecycle related to master data provides a better understanding of the business requirements, ensures that you are getting the right data, and helps to support your projects.*
- **Implement a data stewardship process where possible.** *This provides governance oversight of the data, helps in conflict resolution between groups and confirms data ownership as a business responsibility.*
- **Avoid complex data architecture and integration science projects.** *Focus each release on providing a complete top to bottom slice of scope. This ensures that each project is moving the architecture forward, but still addressing the business objectives.*
- **Make sure you define success and measurements up front.** *Getting agreement on the measures of success and establishing the measurement criteria ensure that value can be proven.*
- **Select a proven approach and proven technologies.** *Experience is the key to making these projects successful.*

Benefits

The creation of an accurate, timely, and rich single view of the customer across channels and lines of business is proving to be a key enabler for reducing costs, managing risk, and increasing revenue and profitability in customer-centric organizations. Companies who have implemented master data initiatives have seen significant returns on their investments.

It has been shown that a five percent increase in customer retention will result in a 75 percent increase in aggregate lifetime profits from an individual customer.⁴ Customer retention and loyalty is just one of the benefits of this approach.

Other benefits include the following:

- **Reduced time to market.** *MDM provides a single system for creating and maintaining product information, promotions, and rich, accurate consumer communications via online and traditional channels. One major retailer used MDM to reduce the time to introduce a new product from four weeks to one day.*
- **Supply chain improvements.** *A single, accurate and agreed upon definition of products and suppliers, made possible by MDM, eliminates duplication, increases buying power, and provides insight into supplier relationships. A major European retailer is combining MDM with WebSphere Portal to allow more than 30,000 suppliers to manage their own master information.*
- **Revenue increases.** *MDM contributes to better customer relationships. By creating a single 360 degree view of customers, MDM helps sales, marketing, and service teams better anticipate customer needs, provide targeted offers, and improve customer service.*
- **Better integration.** *IT departments use MDM to eliminate information silos that have developed across their companies, reducing integration costs, enabling collaboration, and improving business productivity. A major European manufacturer uses MDM to cleanse and synchronize rich, accurate master information across more than 200 instances of SAP and other ERP systems.*
- **Comply with industry mandates and government legislation.** *IBM offers solutions and deep expertise to help customers comply with information centric mandates like Sarbanes-Oxley, The Patriot Act, Basel II, ACORD, HIPAA, and global data synchronization across a broad range of industries.*

MDM also provides common business services to support information-centric procedures across all applications. It gives micro and macro business process capabilities to integrate with or drive business applications. There are always business processes associated with maintaining master information, whether it's setting up new products to be sold, hiring new employees, or eliminating suppliers. The MDM system participates in this process; it can either drive the entire process (the macro case) or be called by another system (the micro case).

Here's a macro example. Before MDM, many business had to enter product information and go through a "new product setup process" in many different internal applications (in a retailer, these applications can number more than 30; in a manufacturer, they can number in the hundreds). IBM's approach to MDM lets customers develop a single composite application that makes it possible to collect all new product information in one place, using a single process, and apply that single master process to ensure the information is complete and validated. Once this is done, MDM software can automatically send the data and automate the individual business processes required to set up the new item in each of those 30 applications.

The micro example is simpler. A particular customer relationship management (CRM) system in a division of a bank will have a workflow for setting up a new customer. Because that CRM system is unaware of the existence of the bank's other systems that interact with customer information, it can simply call out to MDM for a process that will help a customer service rep determine if the customer has any existing relationships with the bank.

Another MDM benefit is accurate business intelligence. Historically, data warehousing initiatives attempted to address data quality problems downstream from applications. Data warehousing never tries to fix the business processes by which inaccurate master data is created in the applications, nor does it try to correct the master data back in the applications. MDM gives businesses a way to correct bad data and the processes that create bad data back at the source applications. Master data management allows accurate objects and structures to be automatically synchronized with data warehouses and analytic applications.

Customer success profiles

Companies that can effectively manage master data can derive much greater returns from their customer, vendor and partner relationships. In order to be successful, organizations need to recognize the master data problem, link resolution to specific and measurable business objectives, and select an approach that maximizes their existing technology componentry and skills.

By selecting a proven approach and proven technologies and implementing best practices in their adoption, companies can achieve very rapid returns on their investments and meet short-term and strategic business goals.

Following is a snapshot of organizations that have leveraged the IBM WebSphere Information Integration portfolio to drive significant benefits around master data integration.

Customer	Project	Business Drivers
JP Morgan Chase	Business Performance Management	Customer profitability
Rockwell Automation	Operational Customer Master	Closed loop marketing
Aetna	Subscriber System of Record	Marketing and customer service
Scotts	Supply Chain Optimization	Reduced inventory costs
DHL Worldwide Express	CRM Rollout	Merger driven consolidation
Celanese	SAP Consolidation	Reduced operating costs, faster time to value
Bombardier	Single View of Parts	More efficient inventory management
Outo Kumpu	Global Parts Master	Eliminate duplication, manage inventory more efficiently
Colonial Electric	Single View of Customer	Consolidation, faster time to value, reduced implementation costs
GMAC Mortgage	Customer / Policy Cross Reference	Cross sell
UK Ministry of Defence	Consolidated Inventory Items & Suppliers	Reduced inventory, time to value, cost to implement
Tyson's	Single View for Global ERP	Acquisition, legacy consolidation, lower costs

Please contact us to learn more about how master data integration using IBM's information integration portfolio can make successful master data management a reality!

IBM WebSphere Information Integration

WebSphere Information Integration solutions provide a broad integration platform that integrates and transforms any data and content to deliver information you can trust for your critical business initiatives. The WebSphere Information Integration platform provides breakthrough productivity, flexibility and performance, so you and your customers and partners have the right information for running and growing your businesses. It helps you understand, cleanse and enhance information, while governing its quality to ultimately provide authoritative information. Integrated across the extended enterprise and delivered when you need it, this consistent, timely and complete information can enrich business processes, enable key contextual insights and inspire confident business decision-making.

For more information

To learn more about master data management, master data integration technologies and IBM information integration solutions, contact your IBM marketing representative or IBM Business Partner or visit:

IBM Master Data Integration:

ibm.ascential.com/solutions/master_data_management.html

IBM WebSphere Product Center: ibm.com/software/integration/wpc

IBM WebSphere Customer Center: www.dwl.com



© Copyright IBM Corporation 2005

IBM Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
11-05
All Rights Reserved

DataStage, IBM, the IBM logo, IMS, the On Demand Business logo, ProfileStage, QualityStage and WebSphere are trademarks of International Business Machines Corporation in the United States, other countries or both.

Other company, product or service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice.

All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only.

¹ A.T. Kearney, Action Plan to Accelerate Trading Partner Electronic Collaboration (Prepared for the G M A - F M I Trading Partner Alliance), 2002

² IBM, Profiling: "Take the first step toward assuring data quality," 2005 (GC-18-9728)

³ IBM, IBM WebSphere QualityStage: Superior Technology Produces Superior Results, 2005 (GC3-18-9744)

⁴ Reichheld, Frederick F. and Thomas Teal. The Loyalty Effect. Boston: Harvard Business School Press. 1996

The IBM home page on the Internet can be found at **ibm.com**