Master Data Management

Making Information Management the Foundation of the Future

William McKnight
McKnight Consulting Group

www.williammcknight.com
More complex and demanding business environments lead to more heterogeneous systems environments. This, in turn, results in requirements to synchronize master data. Master Data Management (MDM) is an essential discipline to get a single, consistent view of an enterprise’s core business entities – customers, products, suppliers, and employees. MDM solutions enable enterprise-wide master data synchronization. Given that effective master data for any subject area requires input from multiple applications and business units, enterprise master data needs a formal management system. Business approval, business process change, and capture of master data at optimal, early points in the data lifecycle are essential to achieving true enterprise master data.

The latest buzz phrase in information management is master data management. It’s yet another take on getting to that lofty goal of a “single version of the truth.” Some will say “Haven’t we been here before?” And you may believe you already do this within the data warehouse or some other existing structure. You could be right. As a matter of fact, a form of master data management is to master it in the data warehouse. However, as information becomes your corporate asset and you wish to control and utilize it as much as possible, these forms of master data management are seldom sufficient. Likewise, the ERP promise of everything in one system leads companies to think master data could be managed there, however, ERP manages just the master data it needs to function and lacks governance rules to change and syndicate master data effectively.
Focus on Subject Areas

While data warehouses focus on all data – and transactional data is the vast majority – emerging master data management programs focus on the high-quality shared data that the organization needs in numerous systems. This data is grouped as “subject areas” and consequently the MDM culture is a subject area culture. Master data will ultimately comprise only 10% - 15% of the volume of organizational data.

One of the main subject areas companies need to get a handle on is customer. It is, by a margin, the subject area with the largest gap between its current state of management and its potential to the organization. Product is another subject area that is a common need. As a matter of fact, in the early days of master data management software, these two subject areas were spawning their own software categories, and consequently, there are quite a few constructs in master data management specific to these subject areas.¹

Other common subject areas that are mastered with master data management are parts, vendors, suppliers, partners, policies, stores/locations, and sales hierarchy. In reality, the list is unbounded and you should let your business needs guide your program’s definition of subject areas and rollout schedule.

¹ You should determine if this type of “intellectual property” around specific subject areas is interesting to your shop. To many shops, it is not because they are not looking for new data, but are looking to harmonize master data throughout the enterprise.
MDM is an iterative project, rolled out across the organization over time. I recommend mastering one subject area at a time, although it is also effective to support one system completely at a time and then move across systems. Often, a combination is best.

Regardless of the rollout strategy, you will want to choose subject areas for MDM that have the following characteristics:

1. High interest to the company
2. High reuse potential in many systems
3. High controversy as to ownership
4. Diverse input to its build
5. Scattered pieces of data throughout the enterprise

These may sound like complicating factors, but that is the point. Without these kinds of data problems, MDM would not be needed. However, it is usually not hard to determine numerous subject areas which can benefit from the MDM value proposition.

<table>
<thead>
<tr>
<th>When</th>
<th>Subject Area</th>
<th>Data Sources</th>
<th>Data Targets</th>
<th>Business Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>2q11</td>
<td>Customer</td>
<td>CRM</td>
<td>Base Customer Information</td>
<td></td>
</tr>
<tr>
<td>3q11</td>
<td>Customer</td>
<td>POS</td>
<td>POS</td>
<td>Transaction Summaries</td>
</tr>
<tr>
<td>3q11</td>
<td>Product</td>
<td>Manufacturing System</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3q11</td>
<td>Product</td>
<td>POS</td>
<td>Replace system of entry</td>
<td></td>
</tr>
<tr>
<td>3q11</td>
<td>Product</td>
<td>Workflow</td>
<td>POS</td>
<td></td>
</tr>
<tr>
<td>4q11</td>
<td>Customer</td>
<td>CRM</td>
<td>Data Warehouse</td>
<td></td>
</tr>
<tr>
<td>4q11</td>
<td>Customer</td>
<td></td>
<td>Data Warehouse</td>
<td></td>
</tr>
<tr>
<td>4q11</td>
<td>Product</td>
<td>Finance System</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1q12</td>
<td>Product</td>
<td></td>
<td>Data Warehouse</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Example MDM Iteration Plan
MDM Challenges

Naturally, every system in the company is utilizing data, and many are using the popular subject areas cited above. Most systems need a customer list to function. Many need a product list. Where and how are they going to get what they need? Obviously, existing systems have come up with a way. However, it’s often a point solution, cluttering the network and, most importantly, inconsistent with corporate standards and with uncertain maintenance. It’s not an efficient, enterprise solution. Master data management is about “build once, use often.”

However, this idea of sharing is not without its challenges and obstacles. Cultures that are decentralized, lack documentation, lack cross-departmental working relationships and lack metadata and a consistent lexicon have challenges. These are not knockout factors, but just factors to be conscious of and overcome.

The benefits of master data management are pronounced and evident and well worth pursuing for most organizations. We are in a perfect storm of:

1. The recognition of information as a corporate asset
2. The need to manage that data operationally
3. The continuing cultural struggles involved in achieving the desired result
What you tangibly get with MDM addresses these factors and includes:

1. A data model for the subject areas, appropriate for enterprise needs – this could be built or adapted from a purchased model
2. A master data publish and subscribe (sources and targets) mechanism and strategy
3. For some master data, workflow processes may support the origination of that data
4. Improved data quality

These deliverables are the result of a variety of disciplines that have emerged over the years in information management. They include data modeling, data integration, data quality and workflow. However, it’s actually the intangible byproducts of these deliverables that may be more impressive. They include the fact that many players in the enterprise will be working from the same lexicon and the same set of data. That is enormously beneficial in improving an enterprise’s processes. The efficient routing of this data means it is brought to bear at the right moment, without delay, such as when the customer is on the phone or very prone to switch carriers after a sixth dropped call this month.
MDM Justification

The most straightforward way to think about the economic payback of master data management is “build once, use often”. Master data must be built for each new system. Systems routinely have up to 50% effort and budget directed towards collecting master data. When master data is built in a scalable, shareable manner, such as within a master data management approach, this will streamline project development time, reducing the time it takes to get new systems up and running. Reducing scope also reduces project risk.

However, having multiple systems working from the same master data is where the ultimate benefit comes from. This is far greater than the total cost of ownership (TCO) “build once” approach, but more difficult to measure. There are efficiencies that come about from elimination of the contention and correlation of numerous “versions of the truth.” One former pre-MDM client used to spend 80% of their “campaign development” time poring through competing product lists to determine which one was the true list for the set of products to be promoted. This left little time for the value-added creativity of the campaign. It also elongated development cycles to the point where time-to-market opportunities would routinely be missed. Clearing up a problem like this is measurable.

I have been speaking of MDM as a support function. That is, MDM is in support of other projects such as campaign management. However, MDM may actually be a prime enabler for many projects such as those centered on customer or product analytics.
What is the value of having customer lifetime value finally calculated? How does it improve campaigns and customer management? Also, there may not truly be a complete set of master data anywhere in the enterprise today – only bits and pieces here and there. MDM may be the mechanism for most effectively introducing master data into the environment, as well as leveraging it into many systems.

You may have a more nuanced situation, but justification will often have to tie back to one of these total cost of ownership or return on investment approaches. MDM actually addresses such a wide range of information and cultural issues that seldom are two business cases alike. I have done business cases focused on the TCO aspects of MDM that will span several projects, on generating customer hierarchies so that customer organizations are understood for risk aversion, on cleaning up customer lists for marketing purposes and on generating customer analytics for more effective marketing, among others.

**MDM Architecture**

There are numerous ways to architect master data for leverage in an enterprise. I have already mentioned the “default” approach for many, which is in the data warehouse. Unfortunately, the (usually) nightly batch-loaded data warehouse is too late in the data lifecycle due to lack of real-time interchange capabilities with the operational environment. Those few who have built “operational” data warehouses that are loaded in real-time and interact with operational systems may have a system that can be
leveraged for MDM. Operational environments must be co-conspirators in building the real-time data warehouse, and few are. Most organizations will need to go beyond the data warehouse.

Sidebar: Whither the Data Warehouse?

MDM does put another chink in the armor of the data warehouse. Many functions which were the hallowed ground of the data warehouse are coming full circle back into the operational environment. Master data management is one such function.

However, by no means does this eliminate the need for a robust data warehouse in the environment. The data warehouse still provides a remedy to the inability to access data in operational environments. It still provides integrated, historical and high-quality data. With MDM in the mix, the data warehouse will actually receive the master data from MDM.

In advanced MDM implementations, since the data warehouse is where the detailed transactions are accessible, analytics can be generated from these transactions and fed back into MDM, augmenting the base data that is there.

Master data can simply be identified where it exists today in the operational environment, pointers to those systems collected in an MDM “hub” and the hub therefore leveraged when master data is needed. For example, the system of record for
the base customer data may be the ERP system. The customer analytics system of record could be the CRM system and the financials related to the customer could be kept in Lawson. These 3 systems would be joined when a full and complete customer master record is needed. Each subject area would have its own strategy.

This “virtual” or “registry” MDM strategy is quickest to deploy because it involves no systemic data movement. However, it often is fraught with performance challenges. My experience with this approach is that it is most effective when limited sets of customer data are needed in the enterprise and data does not need to be moved systemically. To do cross-system joins on the fly for customer, in the example, can be quite costly. And since there is no separate system, data quality is reduced to whatever occurs in the origination systems.

A more common architectural approach to MDM is to physically replicate the master data into a central hub and disperse it from there to other systems that need it. This separate physical hub exists in the architecture as a relational database and is maintained real-time on the inbound side. It maintains its target systems with the data in real-time on the outbound side as well. This approach minimizes network traffic and system complexity. When master data is needed, it will be gathered from the MDM hub. Subscribers do not have to know where the data originated.

There may be other value-added activity occurring at the MDM hub. Various data quality rules could be applied to improve the data. Third-party syndicated data could be
appended to the data at the hub. Workflows could be used to secure business governance to improve the data. Workflows could even be used to completely generate the master data and take the place of an “inbound” system in the architecture. For many implementations, this workflow/governance is the main value proposition for MDM. I will say more about the workflows in the next section.

Master data query is also a function provided by MDM, although it is not as common as master data collection and distribution strategies. For some business functions, all that is needed is query access to the data. Sometimes the implementations generating the analytics within the solution and making them available this way are referred to as Analytic MDM. Whereas this may have been done in an operational system or data warehouse before, when only master data is involved, the MDM hub can be the best place for such a query function. MDM tools provide query “portals” to their databases.

MDM Governance

Selective subject areas will need workflow capabilities to govern the development of a master data record. For example, in order for a new product to be accepted into a retail operation, the Purchasing Manager needs cost, the Marketing Manager needs pictures, the Service Manager needs repair and warranty information and the Training Manager needs features and benefits. Most of this data, or at least the starting points, come from the product vendors themselves.

2 As opposed to Operational MDM, which is focused on sharing consistent and good information throughout the systems.
Some come naturally before the product is ultimately accepted and others come after. Some come before and some come after other critical events in a new product introduction, but all come in a logical sequence of events, the result of which is a master product record.

With MDM, workflow components are used to “pass” the record from one group to the next, perhaps going back and forth several times until the record is complete. The flow does not have to be sequential, either. It seldom is. MDM allows you to design and implement the desired workflow for your master data. This eliminates unwieldy, inconsistent, untimely and inefficient processes.

The workflow defines events, states, transitions and users. The resultant actions from workflow events could be record manipulation or an email or other trigger event for organizational tasks. In the event of a delay, MDM workflow can re-prompt for action and even reassign tasks that are not getting executed. Participants can define their escalation and availability and the routing can adjust accordingly, ensuring (for example) new products continue to be introduced.

MDM Governance can even be backed up to the original entry of the information. In this case of product, that would probably come from the product vendors themselves. This entry can be into a “portal” or another data capture facility that gets the process rolling.
See why cleaning up the inbound side of data with this governance is considered by many to be MDM’s main value proposition?

**Data Quality and MDM**

You could be moving all kinds of interesting data around the organization with MDM, but if it does not adhere to a high standard of quality, it can all be for naught. Actually, that would be an MDM implementation that would not be worth doing at all. Data quality is the absence of intolerable defects. Those defects are defined by the business and include defects in the following categories:

1. Values cross-reference from one entity to another – only customer numbers in the customer master are used in sales transactions for example
2. Uniqueness of identification
3. Expected cardinality of data is adhered to
4. Subtypes and super types perform as expected
5. Values are within their reasonable bounds
6. Conformity of data values to the definition of the field and its data type
7. Consistent formatting as in consistent use of P.O. or PO or Post Office
8. Completeness, or having all mandatory fields contain values
9. Correct values – this is tricky but correct spelling of names, correct addresses, etc.
Consider the many applications supported by MDM and their data needs. How many of them could successfully execute their function with dirty data? Could cross-selling and up-selling be effective if the customers were not unique and complete, or the products were at different levels of granularity or with incorrect attribution? Could you do credit card fraud detection or churn management correctly if you did not have the customer’s transaction pattern correct? Many of these applications have failed in the past because they were not supported with clean, consistent data.

MDM can not only provide these applications with the data they need, it can also share the same data out to all applications.

Of course, the best place to ensure data quality is at the original point of entry. Remediation efforts after that point are more costly and less effective. Many data entry systems allow for free-form data entry, which is a real inhibitor to system success. Now that information is becoming a corporate asset, many companies would like to fix the data and these systems. However, this is not possible in many cases for a variety of reasons.

The workflow processes I spoke of earlier are within MDM and could be a substitute system of entry for the data. Data quality components within MDM could then ensure data quality. Otherwise, when the data does flow to MDM from its point of entry, data quality can be applied at that time and optionally/preferably shared back to the originating system as well as to all subscribing systems.
Syndicated Data

The syndicated data vendor community has existed for a while, but the data has mostly been sourced by organizations with a very specific need, such as a marketing list for a promotion. This is changing as organizations are making the move to the leveragable data store that is MDM. With MDM, organizations have a structure where their efforts in data quality, sourcing syndicated data and organizing a superset of attributes can be leveraged across the organization. Where single applications routinely incorporate data omissions and defects, organizational requirements taken together can cause quality and quantity of data to increase. With MDM, syndicated data has found not only its home in information architecture, but also possibly its “killer app.”

Now that organizations are learning more about how to leverage all manner of customer attributes and analytics, they are increasingly turning to the syndicated marketplace to augment and validate their customer data and create prospect lists. Through a process called a “reverse append,” a small number of fields can go to the vendor and a large number, of varying believability, can be returned.

Syndicated data is valuable and its need has often started MDM programs, since MDM is a central point for the collection and dissemination of master data.
An MDM Project

Given that there are several major value propositions within MDM (publish/subscribe, data quality, workflow, subject area mastery), it follows that no two projects are the same. MDM goals with an MDM project could also be quite divergent. However, there are best practices in methodology that would apply to all implementations.

A top-down methodology involves heavy doses of:

1. Project Management
2. Change Management
3. Time Management
4. Documentation
5. Testing
6. Budgeting
7. Standards
8. Signatures

Top-down in MDM is often fraught with peril. These projects need to move. The extended due diligence yields analysis paralysis and excessive bureaucratic overhead, cost and timelines. Interdepartmental barriers are seldom broken down by top-down approaches. MDM delivery is the best way.
The following words resonate more with a bottom-up, or hybrid, methodology approach:

1. Agility
2. Creativity
3. Open
4. Responsive
5. Ad-hoc accommodations
6. Flexible
7. Reactive
8. Adaptable

MDM can be made to work top-down. However, if you are open to suggestion and offered the opportunity, consider bottom-up for quickest delivery of value with MDM. With the right team in delivery, these approaches deliver high-quality, low-maintenance MDM hubs with minimal breakage and rework necessary. Resources are leveled, problems resolved faster and realistic expectations met better all around.

Regardless of the approach at a high-level, there are unique aspects of an MDM project in addition to the usual phases of planning, design, development and production.

---

3 Talented, multi-capable team members and an organization willing to review milestones instead of project plan items
There is data governance. The team, working with MDM leadership, will establish a master data management strategy to include the following:

1. Data related policies

2. Standards for master data to insure widespread compatibility and security of data

3. Address questions regarding the sharing of master data

4. Identify standards and practices that are consistent with industry practices

5. Establish a publish/subscribe mechanism for master data distribution

6. Build a mechanism to facilitate master data query

The data governance group should be established very early and consists of a cross-section of executives from many areas of the business. This typically includes Marketing, Purchasing, Inventory, Sales, Operations, Finance, etc. Clearly some of the governance activity will be beyond the executives’ ability to directly contribute to, but at least the executives are in place in governance to delegate the responsibility within their organization. It is important to establish support for MDM at this level so that you can turn to data governance whenever a business decision is needed and you can get a quick and accurate response.
One of the unique deliverables of MDM\textsuperscript{4} is data mapping. The mapping exercise will follow the lineage of data from source to MDM to target. Along the way, the data may improve, but also the structure may improve or just be different. Often, the source and target structures are preset. If you cannot share all of the MDM data to the target systems\textsuperscript{5}, at least the unique identification and base data of the subject area must be shared to gain the benefits of MDM.

Agile MDM approaches have a strong element of prototyping. There are both system and user stakeholders for MDM projects. With agile, bottom-up approaches, these stakeholders must understand that their first view of MDM is never the last view and that MDM will evolve over the weeks and quarters, and perhaps years, to grow into what it needs to be to support the enterprise. Systems begin with a basic MDM data model that communicates with target systems in a development environment.

Finally, issues like availability, performance, reliability, scalability, maintenance, usability, connectivity, systems management and disaster recovery are still prevalent in MDM. As a matter of fact, these non-functional requirements are often dictated by the systems that feed data to MDM and rely on MDM for the data.

\footnote{Actually not unique to those who have worked on data integration projects}
\footnote{Sometimes source systems are target systems, too}
MDM Roles and Responsibilities

Roles and responsibilities begin with the business. MDM is not even close to being solely developed by the technical team, especially when workflow is involved. Even when there is light workflow and master data query, leaving existing systems as your primary stakeholders, a project sponsor is needed. This sponsor is ideally a business executive, understands the importance of information to the business, and the importance of MDM to those objectives. The sponsor understands the short- and long-term capabilities of MDM in the organization and actively contributes to shaping them in the form of the iteration plan.

The sponsor will keep MDM out of internal cross-fire and chair recurring executive briefings on the project. The sponsor will chair the aforementioned data governance group and be the executive voice of requirements to vendors in the tool selection process.

Also in the business is the function of data stewardship, which you might think of as the execution arm of the data governance program. These senior business analysts in the subject areas being built into MDM should actually be considered part of the extended development team for the subject area they are the expert in. The functions will include modeling, mapping, data quality rules and workflow design.
The project needs a program manager. The program manager is bottom-line responsible for meeting expectations with the project. To achieve this lofty goal, the responsibilities of the program manager are to:

- Maintain a project plan and obsessively care about the progress on it
- Apply personal skills and judgment to everything on the project
- Match team members’ skills and aspirations as closely as possible to tasks on the plan
- Track all relevant metrics for each iteration
- Meet business objectives with existing or emerging technologies and work on MDM issues with broad technical or strategic implications
- Have significant interface with internal constituents and increase their confidence in the MDM organization
- Ensure necessary documentation is completed and cataloged
- Jump into critical path and challenging technical situations
- Participate in preparation of project documentation
- Represent the project status outside the team

Also on the technical front, there are architects and developers of MDM who will fulfill the requirements using the chosen tools and technologies.
MDM is a discrete deliverable warranting a dedicated team or, at least, corporate governance oversight in order to efficiently meet the requirements of today's organization.

**Selecting MDM Technology**

Also know that to completely meet the MDM challenge, there may be more software required than the core MDM technology. Data quality tools may be needed to address severe data quality challenges. Integration technologies such as an enterprise service bus or some other manner of moving data will be necessary. Some MDM tools do not provide (pre-matching) standardization software. Finally, if you are interested in syndicated data, it is a separate purchase entirely.

Many vendors are claiming the MDM category and the large software companies have clearly positioned themselves in this space. Many buyers are simply extending their product line with their main vendor into the MDM space. This could be workable, but may not always be the best approach. You should base your short list on vendors who seem to support the characteristics of your MDM need, with internal synergy being only one of the criteria. Inevitably, given the breadth of MDM, every vendor will be providing many features that you would not consider paramount.

One of the secrets to successful procurement is to adequately custom value the features of your selected tool to your needs. Salespeople will latch onto various MDM
features that may or may not have direct appeal to your situation. Ultimately you want to make the tool work for you and not the reverse. You do not want to change your requirements to reflect the capabilities of the tool.

After a short-listing process, preferably with an outside consultant, a request-for-information (RFI) is recommended. A shop can probably accommodate two vendors in a proof-of-concept (POC) phase, so a more in-depth request-for-proposal (RFP) is used to whittle the field down to two. Again, use an outside consultant to help you articulate your requirements, short-list, and run the RFI, RFP and POC.

That POC should not be a whiteboard, slide deck review or even a live demonstration. It should be on-site, with your data (scrubbed if appropriate), with actual scenarios that you have developed that represent what you want to do with the tool. The client should have, as much as possible, their hands on the tool in this process. If an army of technicians is supporting the POC with custom coding in the background, you should be aware.
The relevant dimensions of requirements to be developed are:

- Subject areas
- Volume of data
  - Initially
  - Backlog
  - Future expansion
- Source profiles
- Target profiles
- Geographies / languages
- Core problem to be solved
  - Data or subject area
- Data origination mechanism
- Data governance importance
- Workflow importance
- Registry or transactional model
- Operational or analytic model
- Hierarchy management importance
While short-listing, doing the RFI, RFP or POC, the relevant factors you should look for in your tool are:

**Company**
- Viability
- Management
- Sales and support

**Technical**
- Ease of use
- APIs
- Scalability
- Ease of implementation
- Error handling

**Functional**
- Business rule storage
- Data quality handling
- Data access and data storage
- Data integration
- Data governance
- Workflow
- Hierarchy Management

**Culture**
- In-house technology
• Shop personality
• Culture

Usually companies will go into selection expecting and hoping for all factors to be favorable. However, by definition certain factors must be less important than others. In the grand scheme of selection, demonstrated, referenceable scalability that scales well beyond where your shop could ever hope to be, even in the most optimistic case of numerous acquisitions, is not worth pursuing at all costs. Similarly, having 1,000 deployments is interesting, but at a certain, far fewer, number of customers, you can get a sense of commitment and that the tool can do the job.

Furthermore, having many partnerships is nice, but you need only one partner. You need an independent, focused, unbiased, client-centered consulting firm. You need an advocate to the market. Big consulting companies have major conflicts of interest with regards to software selection. I have several times come in behind a large company recommendation to a client in which they recommended a big stack solution for which the firm has resources or where business is done together between the firms otherwise. There’s a place for those solutions, too, but make sure your selection process has the breadth necessary to find the best MDM price-performance for you.

Sometimes less is more – more vendor motivation for your business and your success.
TIBCO Collaborative Information Manager

TIBCO has long been the leader in integration technologies, and their MDM product offering, TIBCO Collaborative Information Manager (CIM), is quickly becoming a pillar in integrating data for leading companies worldwide.

TIBCO is a strong MDM vendor when it comes to harmonizing data across the enterprise in real time. With a focus on providing strong data-modeling capabilities, they are able to support multiple master data subject areas. For MDM governance, CIM provides strong workflow capabilities while maintaining simplicity in the workflow development. Hierarchy management is well developed. Those customer lists that roll up a dozen ways across the enterprise will be well-managed with CIM.

TIBCO as a vendor is engaged and active in the process of demonstrating MDM and supports and welcomes a POC in the competitive process of selecting the MDM vendor.

In Conclusion

Master data management solutions solve real-world problems related to the modern competitive advantage that is information. As organizations wish to harness the information asset, master data management solutions go well beyond anything that has come before. They go into the operational world and harmonize single versions of clean data throughout the enterprise.
CIM represents a cohesive implementation of modern master data management needs. From straightforward midmarket master data requirements to the global enterprise and beyond, CIM is built on the TIBCO foundation that has served the largest and most complex environments the world over for 25 years. By meeting the needs of the modern, competitive organization with a proven master data management solution as well as robust hierarchy management, real-time data movement and strong workflow capabilities, TIBCO is extending that leadership into the midmarket. TIBCO CIM could be your “diamond in the rough” when it comes to MDM tools. It merits a seat at your evaluation.
William McKnight, President of McKnight Consulting Group and very experienced at platform selection, has written more than 150 articles and white papers and given over 150 international keynotes and public seminars. His team’s implementations from both IT and consultant positions have won Best Practices awards.