Seven Tips for Unified Master Data Management

Integrated with Data Quality and Data Governance

By Philip Russom
SEVEN TIPS FOR UNIFIED MASTER DATA MANAGEMENT
Integrated with Data Quality and Data Governance

By Philip Russom
Master data management (MDM) can be practiced many different ways, with various user conventions and a broad array of vendor-built technologies. However, this report focuses on a specific practice called unified MDM. Its seven leading characteristics are:

1. **MDM in the context of a unified program for many data management disciplines.** Unified data management (UDM) is a best practice for coordinating diverse data management disciplines. UDM enables MDM to leverage competency synergies with related disciplines, such as data quality, data integration, and data governance.

2. **MDM as one of many solutions built atop a unified vendor framework supporting many functions for data management.** By using a vendor's unified toolset, developers can share development artifacts (for productivity and consistent standards), plus design solutions that incorporate diverse DM functions. The initial investment in a vendor's unified platform reduces system integration and other costs over time because multiple MDM solutions are built on top of it. A unified platform also accelerates time-to-use for DM projects.

3. **MDM as a series of easily managed projects.** This phased approach avoids risky big-bang projects, and it enables an organization to incrementally grow into multiple MDM solutions that in aggregate amount to enterprise coverage for MDM.

4. **MDM controlled and guided by data governance and data stewardship.** Master and reference data are like all data in that they are subject to the enterprise regulations of governance as well as detailed improvement via data stewardship. A modern, unified platform will provide software functions that automate governance and stewardship tasks.

5. **MDM continuously improved by multiple data quality functions.** Master and reference data benefit strongly from quality measures for standardization, address verification, data enrichment, profiling, monitoring of quality metrics, and so on.

6. **MDM for business people who act as hands-on stewards, not just technical personnel.** A growing number of stewards want and need tool functions designed for them, such as profiling, search, collaboration, and remediation.

7. **MDM organized and optimized via a hub.** Many high-value features of MDM are more broadly disseminated when enabled through a hub, namely collaboration among multiple stakeholders, one-stop governance and stewardship, entity resolution, and publish/subscribe methods.

This TDWI Checklist Report examines these characteristics typical of business programs and technical solutions for unified MDM.

**Definitions of Data Disciplines**

We'll start with basic definitions for some of the data disciplines discussed in this report:

**Master data management (MDM)** is the practice of developing and maintaining consistent definitions of business entities (e.g., customers, products, financials, and partners). MDM's entity definitions and reference data facilitate the accurate sharing of data across the IT systems of multiple departments and possibly outward to business partners. This way, MDM can improve many data-driven initiatives, such as business intelligence, integrating business units via common data, 360-degree views, supply chain efficiency, the compliant use of data, and customer interactions that span multiple touch points.

**Data quality (DQ)** is a family of related data-management techniques and business-quality practices, applied repeatedly over time as the state of quality evolves, to assure that data is accurate, up-to-date, and fit for its intended purpose. The most common data quality techniques are name-and-address cleansing and data standardization. Other techniques include verification, profiling, monitoring, matching, merging, householding, postal standards, geocoding, and data enrichment.

**Data governance (DG)** is the creation and enforcement of policies and procedures for the business use and technical management of data. It is usually the responsibility of an executive-level board, committee, or other organizational structure, although DG is sometimes executed by individuals without a formal organization. Common goals of data governance are to define ownership; improve data's quality; remediate its inconsistencies; share data broadly; leverage its aggregate for competitive advantage; manage change relative to data usage; and comply with internal and external regulations and standards for data usage. The scope of data governance can vary greatly, from the data of a single application to all the data in an organization.

**Data stewardship (DS)** is usually performed by a business manager who knows how data affects the performance of his/her business unit or the enterprise. In addition to daily management responsibilities, a steward collaborates with data management specialists and data governors to direct DM work so it supports business goals and priorities. Many stewards use business-friendly tools to explore and profile data, plus remediate errant or non-compliant data.
For six years running, TDWI has presented an annual conference called the Solution Summit for Master Data, Quality, and Governance. As the name suggests, the conference recognizes that many user organizations practice master data management, data quality, and data governance in a coordinated fashion, along with related practices for data integration and stewardship. Dozens of successful users have spoken at this summit, explaining the many good reasons for coordinating MDM, DQ, DG, and related disciplines:

1. Both DQ and MDM force changes in IT systems and their use. A mature DG program includes a change management process for proposing, approving, and policing such changes.

2. Both DQ and MDM improve data. Many users want to design DM solutions that make multiple improvements in a single solution with a single pass.

3. Master and reference data suffer problems and anomalies, as all data does. They benefit from the improvements provided by data quality functions.

4. Data quality and MDM programs are more sustainable when they demonstrate a positive return for the enterprise. A DG board’s cross-functional mix of people can reveal how DQ and MDM can provide a positive return by aligning with business goals. The simpler practice of data stewardship yields similar results.

5. DQ, MDM, and other DM teams are under pressure to coordinate with other data disciplines. A DG program that focuses on data standards (not just compliance) is an excellent medium for coordination among teams for DQ, MDM, data integration, business intelligence, and data warehousing, among others.

6. As DQ and MDM initiatives grow, they reach further across an enterprise. DG boards enjoy a strong executive mandate that influences the entire enterprise such that DQ and MDM standards have a greater chance of adoption.

From these points, you can see that the coordination of MDM and many other DM disciplines is evolving into a common best practice among business and technology users. The practice has multiple names, such as enterprise data management and enterprise information management (EIM). However, TDWI prefers to call it unified data management (UDM). You can also see that MDM has a prominent place in UDM—especially when coordinated with DQ and DG—and MDM definitely benefits from the coordination.

Defining Unified Data Management (UDM)

From a technology viewpoint, a lack of coordination among data management disciplines leads to redundant staffing and limited developer productivity. Even worse, competing data management solutions can inhibit data’s quality, consistency, standards, scalability, and architecture. From a business viewpoint, data-driven business initiatives (including BI, CRM, and business operations) suffer due to low data quality and incomplete information, inconsistent data definitions, non-compliant data, and uncontrolled data usage.

Forward-looking organizations are addressing these technology and business issues by adopting unified data management, which TDWI Research defines as:

A best practice for coordinating diverse data management disciplines so that data is managed according to enterprisewide goals that promote technical efficiencies and support strategic, data-oriented business goals.

The term UDM seems focused on data management, which suggests that it’s a technical affair. That’s not true because UDM—when performed to its full potential—is actually a unification of both technology practices and business management. For UDM to be considered successful, it should satisfy and balance two requirements:

- **UDM must coordinate diverse data management disciplines.** This is mostly about coordinating the development efforts of data management teams and enabling greater interoperability among their servers. It may also involve the sharing or unifying of technical infrastructure and data architecture components that are relevant to data management. There are different ways to describe the resulting practice, and users who’ve achieved UDM call it a holistic, coordinated, collaborative, integrated, or unified practice. Regardless of the adjective, the point is that UDM practices must be inherently holistic if you’re to improve and leverage data on a broad enterprise scale.

- **UDM must support strategic business objectives.** For this to happen, business managers must first know their business goals, then communicate data-oriented requirements to their managers and to data management professionals. Ideally, the corporate business plan should include requirements and milestones for data management. Although UDM is initially about coordinating data management functions, it should eventually lead to better alignment between data management work and information-driven business goals of the enterprise. When UDM supports strategic business goals, UDM itself becomes strategic.
An organization of any size or sophistication will use multiple tool types for data management simply because there are multiple types of data management tasks, including BI, data quality, data integration, and MDM. Furthermore, the tools employed by users may be from several vendors or may be hand-coded or homegrown. All this diversity can be coordinated at an organizational or team level, but a large or mature UDM program will also need unification at the tool level, which requires that data management tools integrate and interoperate at appropriate points.

Software vendors that produce data management tools have noted users’ need for tight integration and are meeting the demand. For example, several vendors have collected numerous DM tools. The vendor may build or acquire such tools. Either way, DM vendors’ product portfolios have grown in recent years as they fill up with more tools and functions that enable diverse data management tasks.

Furthermore, such vendors continually integrate their DM tools into a unified framework, usually by consolidating most development and administrative functions into a single GUI and by sharing across tools reusable artifacts such as metadata, glossary terms, business rules, profiles, collaborative threads, and data processing logic. The GUI layer aside, the multiple tools of a unified data management platform must also integrate and interoperate deeply in deployment if users are to achieve their primary goal: single, complex DM solutions that embed multiple DM technologies seamlessly, as seen in the earlier discussion of DQ and MDM.

A number of technical users have told TDWI that they would rather use a unified data management platform than take a best-of-breed approach. For example, TDWI’s 2011 next-generation data integration survey asked respondents whether they’re using a DI tool that’s part of an integrated suite of data management tools from one vendor. A mere 9% said yes, but a whopping 42% said they’d prefer to use one.²

TDWI has seen MDM business programs and technology solutions deployed just about every way imaginable, from silos (each focused on a single department, data domain, or application) to fully integrated single solutions for an entire enterprise. Amazingly, each approach succeeds to some degree and coexists with other approaches. The diversity of MDM solution paradigms stems from certain organizational realities:

- Organizational units within an enterprise can have different levels of maturity relative to MDM, which affects their ability to adopt MDM techniques and to integrate with enterprise-scope solutions. For example, one department might first adopt data quality processes to reduce mailing costs or improve claim response times, while another group of departments might be of sufficient maturity (with data governance established, business and IT aligned, and shared services implemented) to build an MDM solution that spans business processes. It’s important that first attempts to increase data management maturity are not disposable and can be leveraged in future phases.

- Organizational units can have varying degrees of interest in integrating data with other units; not everyone considers master data to be an enterprise asset.

- Some MDM solutions provide unique business value on a local, departmental level (for example, customer domain master data optimized for sales).

- Other MDM solutions provide value on an enterprise scale (e.g., product master data, representing every life cycle stage of a product).

- Over time, MDM solutions may evolve from departmental to enterprise in scope, as well as from standalone to integrated solutions. Plus, new MDM solutions are inevitably introduced to address more departments, data domains, and applications.

Given these divergent and evolving business requirements and predilections, organizations are hard pressed to select sustainable methods and platforms for MDM. One path to success is to adopt a UDM platform as the basis for MDM solutions.

A UDM platform offers several technical advantages for MDM:

- In many organizations, the MDM solution landscape is highly diverse, with MDM projects at varying maturity levels and used for various purposes. A UDM platform can tolerate numerous autonomous and unique solutions.

²For more details, see the TDWI Best Practices Report Next Generation Data Integration, available at tdwi.org/bpreports.
• Despite the tolerance for autonomous MDM solutions, the framework of a UDM platform also has all that’s needed to integrate MDM solutions to whatever degree users deem appropriate.

• Users can use a UDM platform to integrate multiple MDM solutions, plus assure consistent data standards across them, using features such as a single data model for all data domains; a business glossary; and numerous shared facilities for metadata, business rules, profiles, processing logic, and interfaces.

• Because MDM tool functions coexist with those for DQ, DG, and DI, a UDM platform simplifies the integration and interoperability of multiple DM tools so users can design and deploy solutions that perform multiple DM functions in a single pass.

• The GUI of a UDM platform accommodates a wide variety of technical and business users, and it empowers cross-functional collaboration over data.

Using a UDM platform for multiple MDM projects also has financial and productivity benefits:

• A unified framework is a foundational investment, which is leveraged financially (and for other benefits) as multiple MDM and other DM solutions are built atop it.

• Leveraging an existing UDM platform can reduce system integration costs for IT as new solutions come online.

• Instead of a risky big-bang approach to enterprise MDM, reduce risk by working into enterprise scope via multiple low-risk MDM phases and projects.

• Subsequent phases and projects can be built on past successes while digging deeper into the platform’s portfolio of features and users’ burgeoning requirements.

• When handled well, unified development fosters the reuse of development artifacts, leading to gains in developer productivity and the consistent governance of data and its standards.

As background, let’s consider the differences between data governance and data stewardship:

• Data governance is the making of policies and standards for governing data’s use and condition on a broad scale, plus translating business goals into data requirements.

• Data stewardship is the pragmatic enforcement of policies for data use and data standards in specific local datasets, plus remediation of errant or non-compliant data.

A DG program gains practical efficiency from incorporating data stewardship. After all, a knowledgeable manager (acting as a data steward) can prioritize data management work, so the work gives the business the biggest bang for its buck. Furthermore, he or she knows the business goals and so can assure alignment between them and work done in data management. In the context of MDM, stewards assigned by a DG program can likewise prioritize and align business with reference data, master datasets, and standards for the exchange and aggregation of reference data.

For these reasons, TDWI recommends that user organizations build data stewardship into their data governance program. In fact, roughly half of user organizations are already doing this in the context of MDM. According to the 2012 TDWI MDM survey, 49% of respondents are using data governance and stewardship functions with MDM today. An additional 41% intend to adopt these functions within three years.1 The survey data corroborates the trend toward unified data management, especially as a combination of DG, DQ, and MDM practices.

MDM requires considerable change, and DG manages change very well. One of the many things that MDM and DQ programs have in common is that both inevitably require changes made to data owned by a variety of departments and sponsors. Similarly, they regularly require changes in how workers use diverse applications. At TDWI, we’ve already seen a track record of success with DQ changes being mandated through the processes and policies of a DG board, then policed and made practical by data stewards. We’re now seeing more organizations do the same with the changes that are required by MDM programs. In fact, TDWI has given awards to organizations that established DG and DS programs before attempting broad-scale DQ and/or MDM.

---

1See the discussion around Figure 17 in the TDWI Best Practices Report Next Generation Master Data Management, available on tdwi.org/bpreports.
Like most enterprise data, master data and reference data present a number of opportunities to leverage and problems to correct. Hence, master and reference data benefit strongly from the many operations available via data quality functions.

Data quality is an important data management discipline because improving the content of data makes it far more valuable in general as a shared enterprise asset. DQ also contributes directly to direct marketing effectiveness, stellar customer service, smoother business operations, and more accurate decision making.

We use the term data quality as if it’s a single monolithic practice, but it’s actually a collection of techniques and tool types, including name-and-address cleansing, data standardization, verification and validation, data enrichment (sometimes called data append or augmentation), and multiple forms of matching, merging, and deduplication. All these have a place in an MDM solution, as do other capabilities we associate with data quality tools, such as quality metrics, business rules, data remediation, data profiling, and data monitoring.

For example, data profiling is often applied to reference data so users (both business and technical) can better understand the state of that data’s problems and opportunities. Likewise, the state of master data is regularly monitored automatically (often measured via quality metrics) to gauge whether it’s fit for purpose. Redundancy is a recurring issue with master and reference data, and matching functions can help with this. When reference data is collected from many sources (and published to many targets), DQ-style standardization is indispensible.

For these and other reasons, in many firms MDM is an outgrowth of a DQ program, sometimes sharing personnel and other resources. The two require many of the same functions for the continuous study and improvement of data’s condition. Furthermore, the two are similar conceptually, and they involve several common skills. That’s because MDM improves master, reference, and other semantic data similar to the way DQ improves physical data. Given the tight integration between DQ and MDM, plus the many tool functions they share, user organizations should consider a vendor’s unified data management platform for the reasons we’ve discussed.
Most of the MDM, DQ, and DG tool features and user practices described as desirable in this report are best done through a centralized hub, as seen in the unified data management platforms described earlier. This is no surprise, given the unified tools, data disciplines, and team members we’ve discussed. After all, it usually takes a central location or software platform to integrate development, interoperate during run time, aggregate data, collaborate among multiple users, and share common data and development resources.

A centralized hub offers a number of benefits for unified data management practices and platforms, especially when applied to MDM, DQ, and DG:

- Tight integration of development environments for MDM, DQ, and DG solutions
- Tight interoperability at run time for the servers of multiple types of DM solutions
- A single data model representing multiple data domains for MDM; the simplicity and consistency this yields is unlikely when organizations deploy multiple, independent MDM solutions on standalone platforms
- Collaboration for many users and stakeholders—both business and technical—that assists with the alignment of MDM and DQ work with business goals
- Fewer conduits for data movement and solution development, which simplifies governance and stewardship for MDM and DQ
- Broad (perhaps enterprise-scale) matching, entity resolution, de-duplication, and best-record matching
- Cross-data-discipline data flows, workflows, process designs, solution designs
- Development assets shared across multiple data disciplines, for greater developer productivity and consistent standards (such as, shared metadata, business glossary, profiles, quality metrics, business rules, transformation logic, and processing methods)
- Aggregated, improved, governed, and secured master data, possibly in real time via services
- Rich collections of interfaces, services, and interoperability options, all shared and controlled through a single central unified framework
- One point of administration for multi-data-discipline solutions
ABOUT OUR SPONSOR

www.sas.com

About SAS® Data Management
SAS is a recognized leader in data management and business analytics software and services. SAS master data management (MDM) employs “Phased MDM” to help customers bridge corporate silos, align business and IT, and drive a consistent, accurate view of their data. Core differentiators include:

Embedded Data Quality—SAS pioneered MDM solutions built on top of a market-leading embedded data quality platform.

Agile Foundation—SAS MDM can be more rapidly deployed at reduced services integration costs, and integrate with more sources because of its embedded Data Management Platform underpinning.

Unique “Phased MDM” Approach—Start with a data quality or data governance challenge, grow into a single-domain, batch-fed MDM project, and then migrate up to an enterprise MDM deployment when you are ready, leveraging existing investment by building on the same data management foundation.

Unified Framework—SAS Data Management leverages a microservice architecture to reuse shared services across its entire portfolio, only deploying when needed. The SAS Data Management Console is a simplified common user interface that exposes users to data management capabilities including entity search, workflow, process orchestration, job monitoring, and data issue remediation.

Pervasive Data Governance—Embedded data stewardship, reference data management, business glossary, data quality monitoring, and data remediation capabilities to improve business and IT collaboration.

Data Management Consulting—SAS provides deep consulting expertise to support both the data governance and software delivery components of deploying MDM solutions.

For more information about SAS Data Management software and services, visit sas.com/data.

About SAS
SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 70,000 sites improve performance and deliver value by making better decisions faster. Since 1976 SAS has been giving customers around the world THE POWER TO KNOW®.

ABOUT THE AUTHOR

Philip Russom is the research director for data management at The Data Warehousing Institute (TDWI), where he oversees many of TDWI’s research-oriented publications, services, and events. He’s been an industry analyst at Forrester Research and Giga Information Group, where he researched, wrote, spoke, and consulted about BI issues. Before that, Russom worked in technical and marketing positions for various database vendors. Over the years, Russom has produced over 500 publications and speeches. You can reach him at prussom@tdwi.org.

ABOUT TDWI RESEARCH

TDWI Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence and data warehousing solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

ABOUT THE TDWI CHECKLIST REPORT SERIES

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.