

Paper 125-2012

Evolving from Data Management to Master Data Management

Wilbram Hazejager, DataFlux Corporation, Heidelberg, Germany

ABSTRACT

The proliferation of enterprise-level applications (along with expectation for shared, synchronized information) drives the need for the development of a single view of the key data entities in common use across the organization. Therefore, master data management (MDM) has become one of the fastest-growing application areas in recent times. This presentation provides an overview of how DataFlux® qMDM can enhance your data management initiatives and put you on the path to developing a complete view of your data.

INTRODUCTION

The increasingly growing volume of data coming into an organization from internal and external sources, and the spread of data across the enterprise causes many companies to have nightmares. There is no “truth” to the data that they hold on customers, products, inventory, assets, or even employees. And it is becoming more difficult to make sound decisions (both strategic and operational) about the business.

For the past 20 or 30 years, companies have invested in applications that are designed to maximize the data they have, hoping to provide a structure for data on an ongoing basis. These business applications are often designed to augment or administer a business process or processes (for example, a customer relationship management (CRM) system manages sales and marketing activities). From that standpoint, much of the work on these applications has focused on just that – the processes that feed into a “data supply chain.”

During the 90s, data warehousing became popular and provided a more structured approach to integrate data from those business applications and store the data centrally using consistent definitions to enable BI and analytical reporting to support business decisions. The data in the source systems remained unchanged. Bringing the data together just by itself was already a major achievement. There were still lots of legacy and proprietary systems around that often made it difficult to just get to the data. Sharing data throughout an organization that was traditionally only visible in one part of the business was initially met with suspicion and sometimes even reluctance. During the 90s, major strides were made in terms of industry standards for data access and integration of application functionality. You could say that at least from a technology perspective, data warehousing paved some of the way for the next step in data management.

Master data management (MDM) arrived by the early 2000s as the panacea for the problems of disparate, disjointed data. The spread of data across different business lines coincided with the adoption of new regulatory measures, such as Sarbanes-Oxley and more recently Solvency II, designed to mandate a higher level of rigor for the integrity of enterprise data. Organizations were quick to adopt MDM technologies. It should be noted though that MDM is a discipline that incorporates the people, processes, technologies, and methodologies for creating an authoritative view of the enterprise. The results from MDM implementations have been underwhelming, if not frightening. There are many reasons why MDM projects fail.

In this paper we will show you an evolutionary approach to MDM that starts by looking at the data management infrastructure in context of data governance maturity, before looking at specific DataFlux technologies to apply to support MDM projects in the various stages of data governance maturity. Such an approach minimizes risks in your MDM projects.

EVOLUTION FROM DATA MANAGEMENT TO MDM

The data management challenges facing today’s businesses stem from the way that IT systems have evolved. Enterprise data is frequently held in disparate applications across multiple departments and geographies. The confusion caused by this disjointed network of applications leads to poor customer service, redundant marketing campaigns, inaccurate product shipments, and, ultimately, a higher cost of doing business.

To address the spread of data – and to eliminate silos of corporate information – many companies implement enterprise-wide data governance programs that attempt to codify and enforce best practices for data management across the organization. Although the goal is clear – the quality of information *must* improve to support core business initiatives – there is no definitive roadmap for starting these projects.

For any organization, the first step to address the quality and value of corporate data is to take an honest assessment of the data management infrastructure. Through the Data Governance Maturity Model, organizations can identify and

Evolving from Data Management to Master Data Management, continued

quantify precisely where they are – and where they can go – to create an environment that can deliver and sustain high-quality information.

In the past, business units were concerned only with entering and tracking data to meet the needs of their specific departments. The result for the enterprise was a buildup of redundant, inconsistent, and often contradictory data, housed in isolated departmental applications from one end of the organization to another.

The impact of poor-quality data was initially felt in applications that focus on customer information: database marketing, data warehousing, and customer relationship management (CRM). Data quality and data integration technologies, when applied locally, could address some of these problems within each application. However, this only led to silos of consistent, accurate, and reliable data. The goal is to move beyond those silos and find a way to manage this data across departments, applications, business units, and divisions.

Master data management technology will help to create a more unified view of the enterprise, so that managers and executives can create strategies that make the company more profitable.

Stepping up from applications on business unit level to cross-department and enterprise-wide, does not only mean adopting new applications and technology, but often more importantly requires a different approach to process and people, ultimately leading to higher benefits for the organization.

The following diagram summarizes this approach.

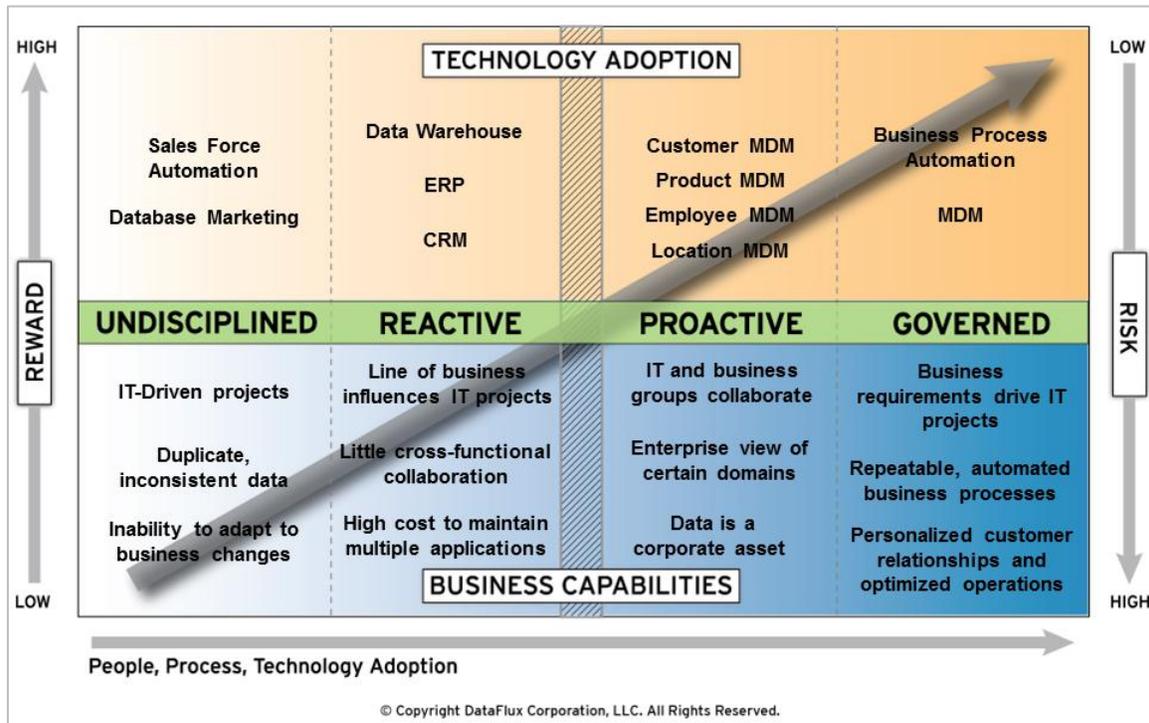


Figure 1. DataFlux Data Governance Maturity Model

Now focusing more on the data quality and data integration needs that are associated with the various levels in the Data Governance Maturity Model, you could characterize these as follows:

Evolving from Data Management to Master Data Management, continued

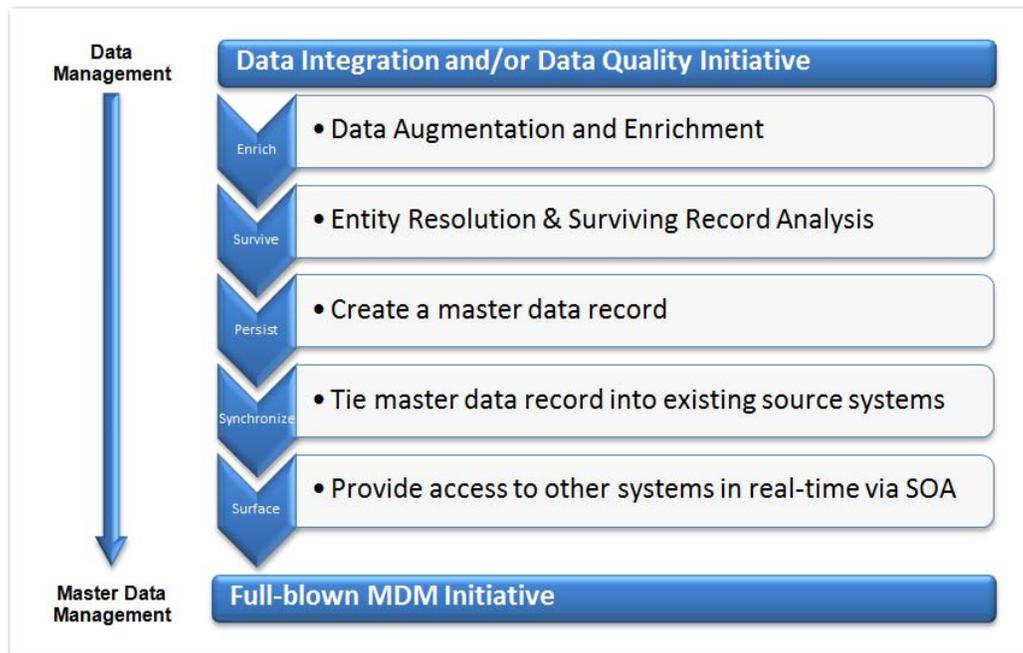


Figure 2. Evolving from Data Management to MDM

The movement from data management to MDM truly is an evolution. The process commonly starts with basic data management capabilities, such as data standardization or data matching. The next step could be data augmentation using the address verification databases, or the identification or gender analysis functionality. The data might then go through some sort of segmentation and analytics prior to creating a “master data file.”

Once the “master data file” is in place, users throughout the enterprise will begin to see the benefits and opportunities generated by this single corporate view of the data. They will seek to use the master data records within their existing operational and transactional systems. This will eventually evolve into the desire to have the data available to other systems in real time through Web services. At this point in the process, the customer has evolved to a full-blown MDM initiative.

DATAFLUX SOLUTIONS THAT HELP EVOLVE FROM DATA MANAGEMENT TO MDM

DataFlux provides a portfolio of integrated products to support any of the following: data quality, data integration, and MDM initiatives.

DataFlux Data Management Studio: Entity Resolution

DataFlux Data Management Studio provides a single interface for both business and IT users to plan, implement, and monitor the rules to manage data throughout the organization. Part of the DataFlux Data Management Platform, DataFlux Data Management Studio provides a unified development and delivery environment, giving organizations a single interface to analyze, improve, and control data and drive enterprise data management.

DataFlux Data Management Studio offers a unique set of workflow tools built on an industry-leading technology platform that encompasses every facet of the data management process, including data profiling, data access, data transformation, data enrichment, data monitoring, and entity resolution.

Entity resolution allows you to:

- identify individuals across multiple data sources from incomplete and non-obvious relationships
- intelligently manage entity resolution routines through advanced fuzzy-matching technology
- create multi-record clusters, confidence scores, and scatter plots to determine potential clusters
- analyze the suitability of data elements as potential identifying attributes
- recognize when slight variations suggest a connection between records

Evolving from Data Management to Master Data Management, continued

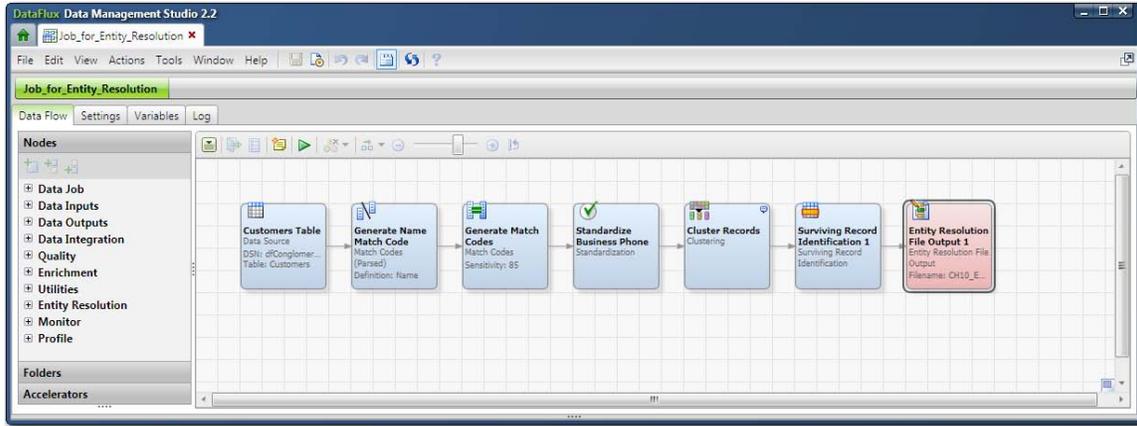


Figure 3. Screenshot from DataFlux Data Management Studio – Job to Create Entity Resolution File

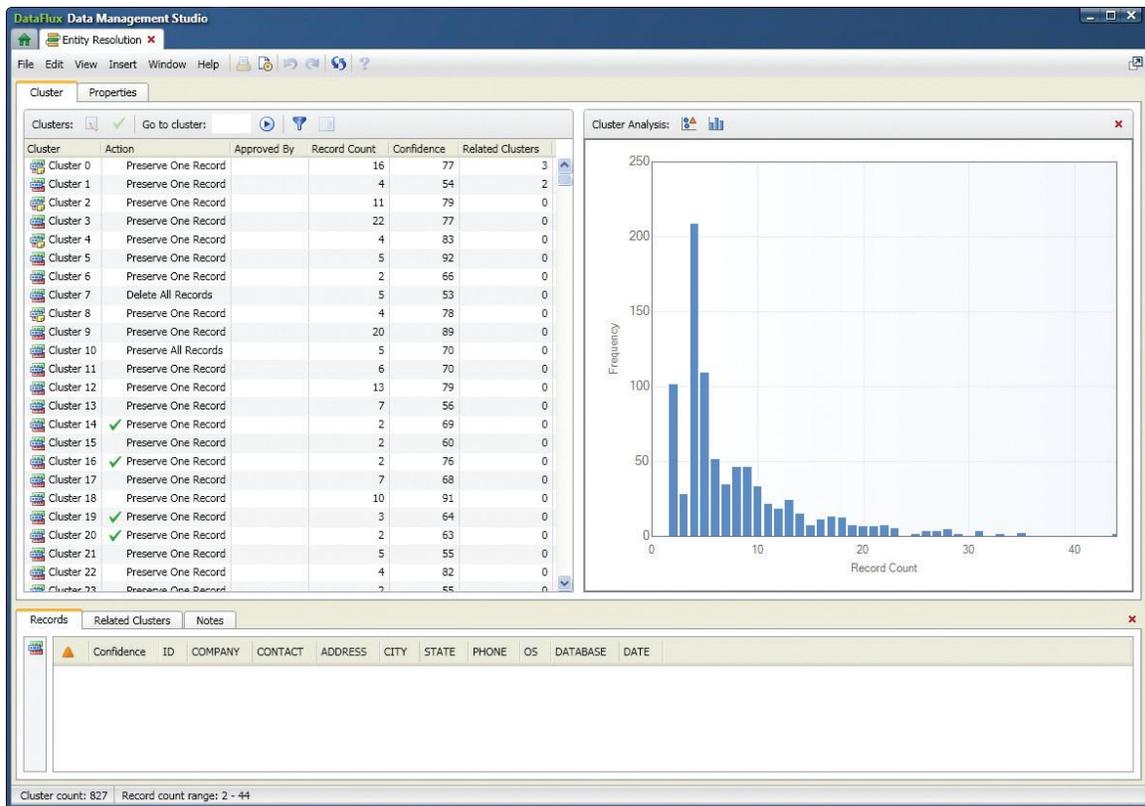


Figure 4. Screenshot from DataFlux Data Management Studio – Entity Resolution Editor

This technology has been applied over the years in many data quality projects. As a result, DataFlux has created Master Data Management Foundations, which incorporates best practices gathered during those projects.

DataFlux Data Management Studio: Master Data Management Foundations

MDM efforts are often part of a larger data management or data governance initiative within the organization. With Master Data Management Foundations, you can configure a pre-built master data model within the DataFlux Data

Evolving from Data Management to Master Data Management, continued

Management Studio application, providing the foundation for accelerated MDM deployments and seamlessly move to DataFlux qMDM for advanced MDM capabilities.

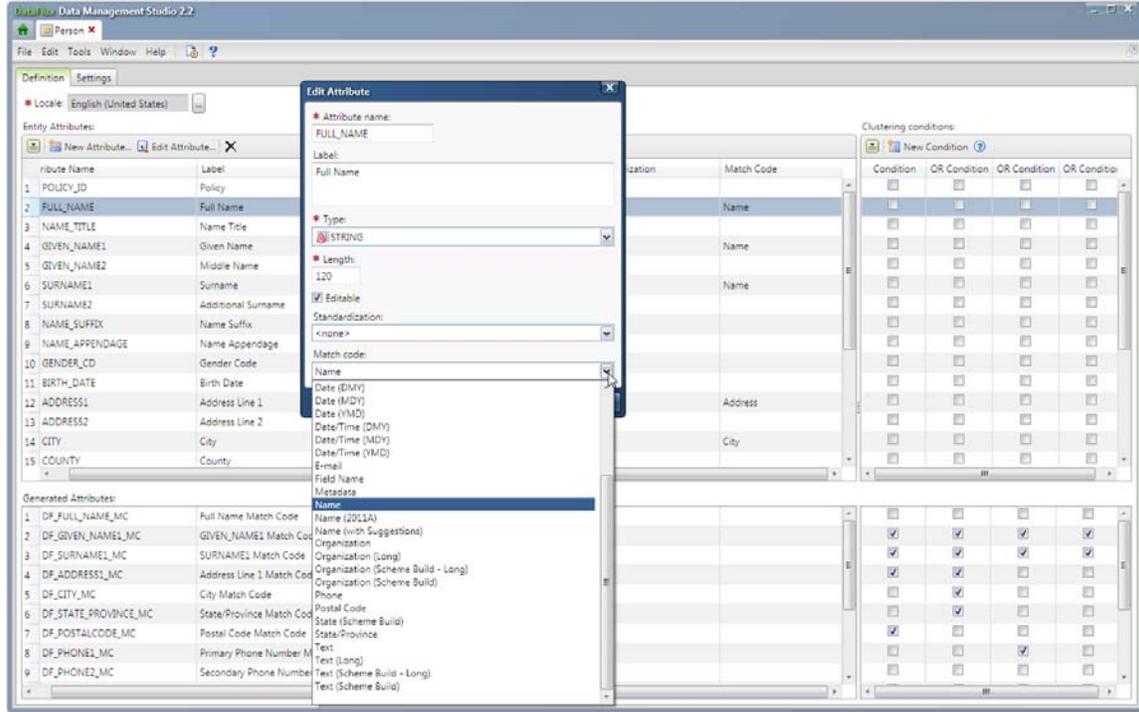


Figure 5. Screenshot from DataFlux Master Data Management Foundations – Configuring a Pre-built Master Data Management Model

With DataFlux Master Data Management Foundations, you can do the following:

- integrate the creation and management of master data resources with comprehensive data management practices
- create a hub of master data based on a subset of your existing data through a phased approach
- combine MDM capabilities with matching, clustering, and other data management initiatives
- conduct batch processing with an architecture that supports many MDM implementations without unnecessary complexity
- connect to MDM hubs as if they were any other data target

DataFlux qMDM

DataFlux qMDM leverages the power of the DataFlux Data Management Platform, which provides award-winning solutions that are designed to help organizations increase the value of their data assets. DataFlux qMDM allows for the integration of results created using DataFlux Master Data Management Foundations.

With this foundation, DataFlux qMDM supports end-to-end data governance projects by providing a wide range of data governance technologies, including the following:

- a data stewardship console that allows data stewards and business analysts to analyze existing data resources and design MDM workflows and business rules
- role-based access for data stewards that controls user access to critical master data and protects sensitive and confidential information such as Social Security numbers and other personally identifiable information
- auditable, historical views of master records, meaning that the relevant data history that is required for governance is maintained in the solution infrastructure

Evolving from Data Management to Master Data Management, continued

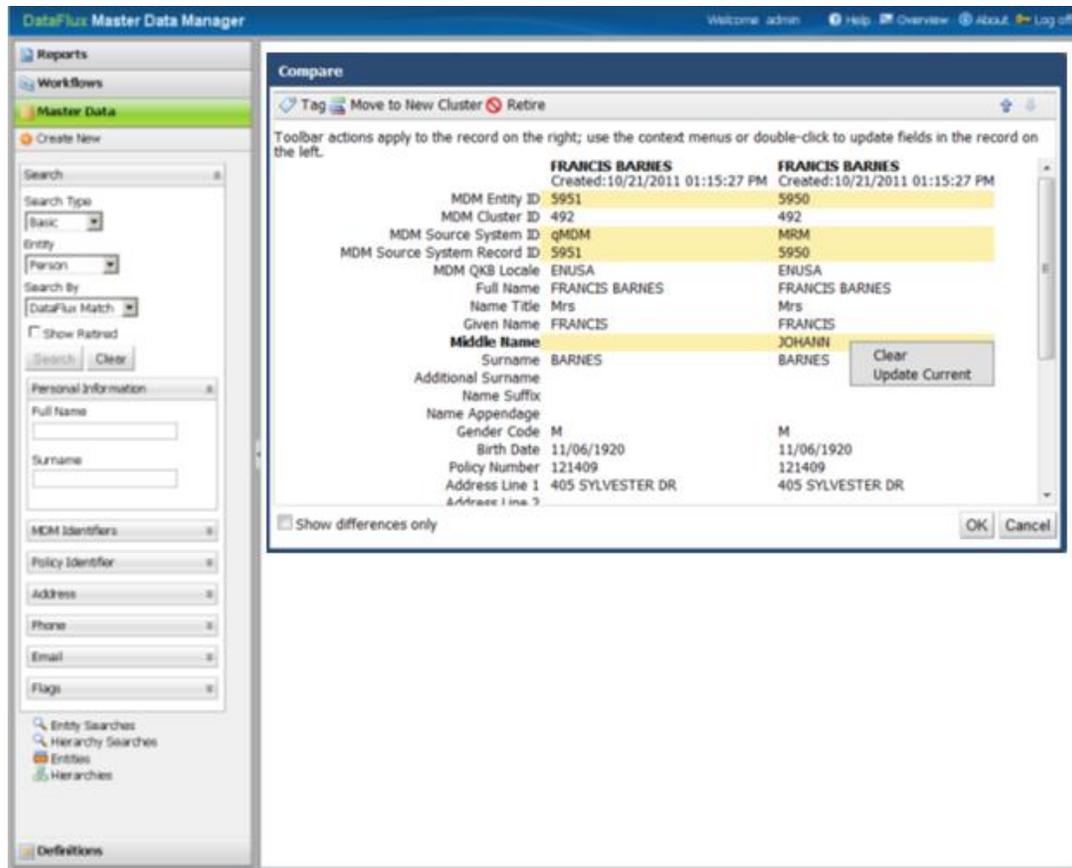


Figure 6. Screenshot DataFlux qMDM Master Data Manager

DataFlux qMDM has the following features and functionalities:

- **Multi-domain data model** – Create master data domains for party (citizen, customer, student, and so on), organization, site, supplier, product, asset, or other data elements
- **Metadata discovery and data profiling** – Begin the MDM deployment with a solid understanding of the health of incoming data – and jump-start your MDM project
- **Entity resolution and matching** – Engage a patented data matching engine to create and maintain a master record
- **Data stewardship** – Control the master data hub, as well as the processes and workflows that govern the creation of master data, through an intuitive, business-focused interface
- **Data quality improvement** – Use DataFlux best-of-breed data quality technology to standardize and rationalize data during the MDM data lifecycle
- **Hierarchy management** – Manage inter-related data elements, such as tiered and networked hierarchical structures
- **Survivorship** – Validate master data representations through entity definition, resolution, best record selection and editing, and the creation of a universal identifier
- **Data integration** – Engage batch and real-time data integration processes to manage the master hub repository
- **Reporting** – Monitor the master repository and the MDM process through interactive dashboards and scorecards
- **Data lifecycle management** – Manage the entire data management process, including data loading, data updating and data management
- **Entity lifecycle management** – Establish review cycles before changes are made live, route tasks to people or roles for completion, and monitor the state of the workflow as it progresses

Evolving from Data Management to Master Data Management, continued

DataFlux Solutions Compared

In the previous sections we have explained the various DataFlux solutions for each step of the way in your evolution from data management to master data management. The following diagram provides a high-level comparison of functionality in each of those solutions.



Figure 7. Evolving from Data Management to MDM – DataFlux Solutions Compared

Remember that both DataFlux Master Data Management Foundations and DataFlux qMDM are built on top of the DataFlux Data Management Platform. This allows you to re-use jobs that were built using the DataFlux Data Management Studio and also allows you to use DataFlux Data Management Studio to enhance the type of processing that takes place in Master Data Management Foundations and DataFlux qMDM. Furthermore, you can take definitions used in Master Data Management Foundations and re-use them in DataFlux qMDM.

CONCLUSION

The amount and the complexity of corporate data in every business are growing. Data is increasingly shared across corporate and geographical boundaries. And the success of any organization will ultimately hinge on the ability to maintain a coherent view of data, both now and in the future.

For any company that wants to improve the quality of its data, it is critical to understand that achieving the highest level of data management is an evolutionary process.

DataFlux provides software for every step of the way in your evolution from data management to master data management. Using pre-built assets reduces risk and allows you to provide results and value quicker. Using a tightly integrated solution lowers integration costs. All of this combined ultimately leading to lower total cost of ownership.

DataFlux Master Data Foundations and DataFlux qMDM leverage the award-winning DataFlux Data Management Platform, which provides the data quality, data integration and data governance framework that drives successful MDM programs.

Evolving from Data Management to Master Data Management, continued

REFERENCES

- DataFlux Corporation. 2011. A DataFlux white paper. "The DataFlux Approach – The Practical Value of MDM." <http://www.dataflux.com/Knowledge-Center/White-Papers.aspx>.
- DataFlux Corporation. 2009. A DataFlux white paper. "The Data Governance Maturity Model – Establishing the People, Policies and Technology That Manage Enterprise Data." <http://www.dataflux.com/Knowledge-Center/White-Papers.aspx>.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Wilbram Hazejager
DataFlux Corporation
Heidelberg, Germany
Work Phone: +49-6221-4150
E-mail: Wilbram.Hazejager@dataflux.com
Web: www.dataflux.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.