Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2011 Proceedings - All Submissions

8-6-2011

Analytical Master Data Management 2.0

Stephan Zoder, MIPP
IBM, szoder@us.ibm.com

Analytical Master Data Management 2.0

Stephan Zoder, MIPP
IBM
szoder@us.ibm.com

April 2011

Abstract

Analytical Master Data Management (aka aMDM) has been an on-again, off-again topic in the information management space. Some describe it as just a use case within MDM, others limit it to dimension management for BI purposes. For some aMDM appears to be a solution looking for a problem. This paper will discuss some concepts of how a next generation version (2.0) of aMDM should evolve and be implemented from its current status quo "Dimension/Hierarchy Management" (1.0) serving data warehouses. aMDM 2.0 creates, manages and serves not only hierarchical meta data but also real-time analytical facts to BI and OLTP applications leveraging existing opMDM functions to ultimately support business transactions. The following analysis is grounded in functional scenarios by industry supporting the evolving usage patterns; opMDM, aMDM 1.0, aMDM 2.0. Targeted readers are: IT managers, CIOs, enterprise architects, business analysts and line-of-business (LoB) leaders. It is the author's goal to prove how the value of linking transactional and analytical environments. The research uses a Delphi-style analysis via solution design and data governance workshops in a number of industry settings. All interviewees hold thought leadership positions in their respective market-leading organizations and fill similar positions as the aforementioned target audience of this paper.

KEY WORDS

Analytical Master Data Management, aMDM, MDM for Business Intelligence, Data Warehouse, Dimension Management, Hierarchy Management, Business Analytics

INTRODUCTION

> "I will build a car for the great multitude. It will be large enough for the family, but small enough for the individual to run and care for. It will be constructed of the best materials, by the best men to be hired, after the simplest designs that modern engineering can devise. But it will be so low in price that no man making a good salary will be unable to own one—and enjoy with his family the blessing of hours of pleasure in God's great open spaces. (English, 2009)
>
> - Henry Ford

Considering the given technology, resource constraints and knowledge at the time, this was a lofty goal; to be everything to everybody at a low price. Nevertheless, over 100 years later we attribute Henry Ford with spearheading the mass production of passenger vehicles with the Model T? What he did not mention in this quote is that he took an already existing set of technologies; the concept of the modern steam engine, wheels, the internal combustion engine, adding low frills engineering and conveyor belt mass production. Overlooking the simplification here – you had the grandfather of the modern automobile and its industrial manufacturing process. Ingenious people continued to enhance it to increase not only its efficiency, safety and comfort by introducing gadgets like fuel injection, ABS, cruise control, day time running lights, air bags, navigation system or even - dare I say it - the cup holder, but they also added additional hardware to effectively extend its use. One such item was the automobile-attached snowplow attributed to David Munson (1911). The snowplow allowed a vehicle to go from moving passengers and cargo from A to B to doing it in three feet of snow effectively making transportation and travel and by extension tourism and commerce to remote areas a year-round affair. Moreover, one such relatively simple hardware item on the vehicle upfront allows following vehicles to leverage its capability. Snowplows staggered in sequence can clear a highway in no time.

1+1=3!

Imagine a similar concept, i.e. the snowplow for a Model T, for your ever growing key data in the enterprise to enhance your forward vision based on the past and ability to operate effectively? Would it not be great to achieve this by just extending the reach of an existing key information technology, say an instance where you maintain your most precious corporate data, the data warehouse, the ERP system or the billing system, effectively adding more value to more people's jobs by leveraging this legacy investment. Well, the Master Data Management for Business Intelligence concept (MDM for BI), aka Analytical MDM (aMDM) is exactly that extension. It combines traditional opMDM and data warehouse capabilities to exponentially increase their individual value-add. At the very least, it establishes trust in reported numbers by centrally organizing your data structures, aka taxonomies, for auditable consumption by many downstream BI systems and users depending on their particular needs. (White, 2009) You may say that this is nothing new as such. Gartner already labels this capability as Analytical Master Data Management; others dub it Dimension Management. Ideally: however, the "classic" aMDM – if there is such a thing - should also be combined with and leverage transactional-oriented capabilities native to the opMDM paradigm, no matter if data exists in a physical or virtualized (registry, federated) sense. In this case, key master and metadata useful to two or more applications, processes or departments is not only structured for initial BI use but also cleaned, persisted and managed to support any business transaction with a single, trusted, full operational view of a customer, a supplier, a product, an asset or an account. This "new" data can now be used to automatically drive end-to-end business processes starting from business insight to optimized transactional execution as it adheres to enterprise-wide data definitions and relationships of supporting data around customer, product, asset, etc.

AMDM 2.0 DEFINED

How does this aMDM instance leverage and enhance warehouse data, previously assumed to be master data just like ERP data a decade ago? It does so by not only using the same hierarchies for OLTP (online transaction processing) and BI consumption but also by publishing upon request aggregates or calculations established at runtime via federated query logic of historical data from an active data warehouse. (Hechler, 2011) To optimize performance and in order not to abuse native data warehouse capabilities and constraints I propose also the addition of what some call a "write-back" function, i.e. the persistence of aggregates and their frequently refresh for OLTP system consumption, say in form of a "quick KPI view" of a customer with less fancy SQL and graphing, and thus "cheaper" trending. The resulting combination of operational (customer name, address, etc.) and analytical data (hierarchical metadata and KPIs) allows for a more effective joint use of the two data types and IT environments – operational and analytical. (*Figure 1*) This approach represents the most comprehensive, trusted "Single View of x" to feed key business processes. (Hayler, 2010) Here x represents any data domain of interest.

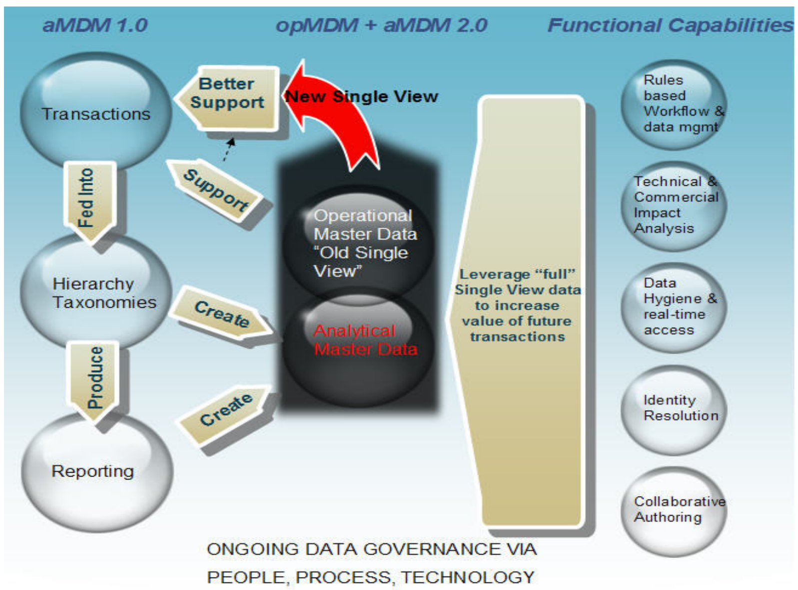

Figure 1. Analytical MDM – What is it?

AMDM AND THE LIMIT AND REACH OF HIERARCHY MANAGEMENT

While dimension management is a key feature or aMDM and key for some organizations, many others see it as a "frills" problem or something easily achieved by consolidation, data governance policies, business rules often tackled via a low-key custom in-house development. While there is great value in it by improving data governance, I believe that its scope falls dramatically short of a more holistic, end-to-end vision, which drives automated decisions where needed. Moreover, what good is it if the sources of the underlying data to be analyzed work of a different hierarchy which is defined, manipulated and governed differently than the one(s) used by the data warehouse. This creates operational inefficiencies, data quality issues and much more. As such, the new operational twist of the aMDM concept can add or ignore hierarchy management capabilities. Hierarchy management only becomes a relevant question if multiple warehouse/dimensional environments are accessed to deal with the pressing BI requests and if hierarchies are highly dynamic. More importantly, uncomplicated user environments should allow business analysts to assemble data and logic required for a good decision instead of developers and data administrators, severing their dependence from a backlogged IT department, as we know it from the data warehouse world. Analysts should also be able to foresee the downstream impact of considered changes by adding required data elements (calculated or not), visualizing past changes and simulating outcomes at a moment's notice. (White, 2007) This requires a new paradigm about the way you think about your data. Now, data is at least logically situated in one location, massaged, stored and exposed in many different ways supporting transactions and BI alike by adding a new product, offering a customer a discount or ignoring a restrictive return policy as well as a common compliance reporting taxonomy. As transactions change the facts about a customer, product, etc. their hierarchical position changes together with their transactional history. Data previously thought of at-rest is now in-motion and depending on the frequency of updates could very well be considered in-flux. (*Figure 2*)

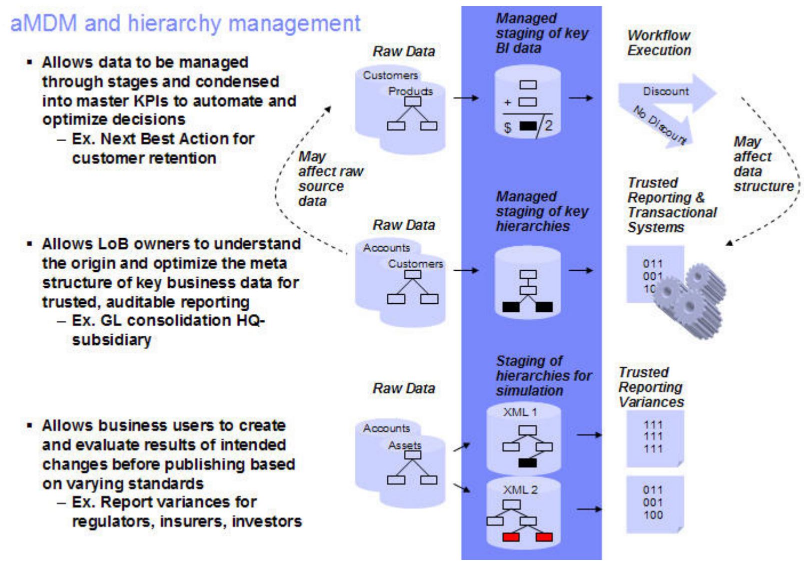

Figure 2. The power of hierarchy management in aMDM 2.0

If we take a closer look at what this vision looks like you will agree that it is the only true way of solving the "trust in information" and "BI as decision driver" challenges. The components required to achieve this are:

- Centrally persisted, aka consolidate approach, or virtualized, aka consolidate & propagate approach, instance of master data used in reporting
- Comprehensive means to clean (standardize, dedup) and structure master data near the source of entry (ideally at time of entry) for analytical and transactional consumption
- Ability to link and resolve multiple hierarchies of multiple data domains at every level
- Stage data to support hierarchies based on external or internal taxonomies
- Provide data lineage information on an attribute level
- Temporarily sandbox changes to process and data structures
- Relate financial data to sandboxed changes to simulate and assess bottom-line impact
- Assess how many downstream systems, ETL jobs, taxonomies and services are affected and rank them by priority (quarterly regulatory reports, number of users, etc.)
- Store workflow results of approved changes in a comprehensive way
- Automated or manual push/pull of changes to reporting and transactional systems
- Workflow support to manage process tasks by priority and impact
- Frequent capture and aggregation/calculation of key business insight (aka Key Performance Indicators and predictors or KPIs and KPPs) to drive business decisions supporting transactions
- Scalable accessibility of captured master data to third party process automation and advanced analytical systems

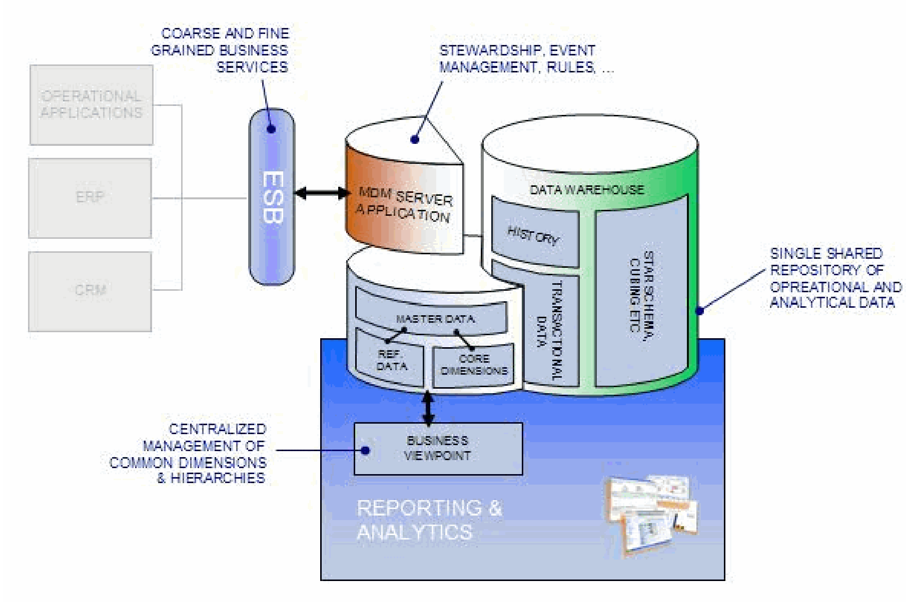

Figure 3. Sample illustration of topology - Data warehouse centric view (Byrne, 2010)

BI MASTER DATA IN ACTION

The following sections illustrate how different business functions approach master data in an analytical sense to gain new insight and drive bottom line results. Industry use case examples further depict how a particular sector deals with these key requirements and reaps the benefits. Graphical depictions of the iteratively evolving concept illustrate how increasing integration with additional capabilities creates a path to grow with the business need.

PRODUCT MANAGEMENT

As a design-time function in a business, product management is key in understanding what products and services to launch, alter and withdraw next. Ideally, product managers should collaborate and minimize configuration work, shortening the time from idea to cash, while understanding the financial impact of making such changes before committing them to the operational world.

A Financial Industry Example

The importance of a complete solution becomes clear when we look at an example of how a financial institution assesses and manages the introduction of a new product today. A sales leader gets a new sales goal assigned and looks for options on how to achieve it. He can try to do more with the same (staff, products, customers) or more with (a little) more – say a new credit card product. Banks typically have very few core products but a lot of variances within each. He will likely request product management to come up with a few new product ideas on how to twist an existing product to tap an existing or new customer base. One option may be a student card with no annual fee, minimum purchase requirement, a high interest rate and a free-concert ticket benefit. As such it is key that any new product or service for operational as well as reporting use exists only once (enterprise catalog) within the organization and can be, if needed, easily linked to instances for particular uses, e.g. sales, fees, interest, deposits, loans, cards, etc. Even more important is that a team of experts can collaboratively create, review, change and release the new product by re-using existing product structures while business rules validate applicability. Qualitative reviews here are very important as regulatory compliance can minimize compliance risk. Product information, hierarchy management and life cycle management capabilities within an MDM solution bring key functionality to bear in this case.

A Retail Industry Example

A retailer will act similarly by conceiving a new twist on an existing product, possibly just changing a bottle size, the color, the ingredient, the fabric, the offered locations or position in the store. What happens next is a very cumbersome, very manual process involving the creation of a new product hierarchy with some new attributes in the catalog, which then need to be pushed out to or synchronized with a variety of analytical systems to ensure a common BI experience and trusted results. What could make this even more interesting is a global roll-out which needs to address local packaging or color schemes. *(Figure 4)* The receiving systems could be formulation, ERP, real estate, store placement, risk and many others.

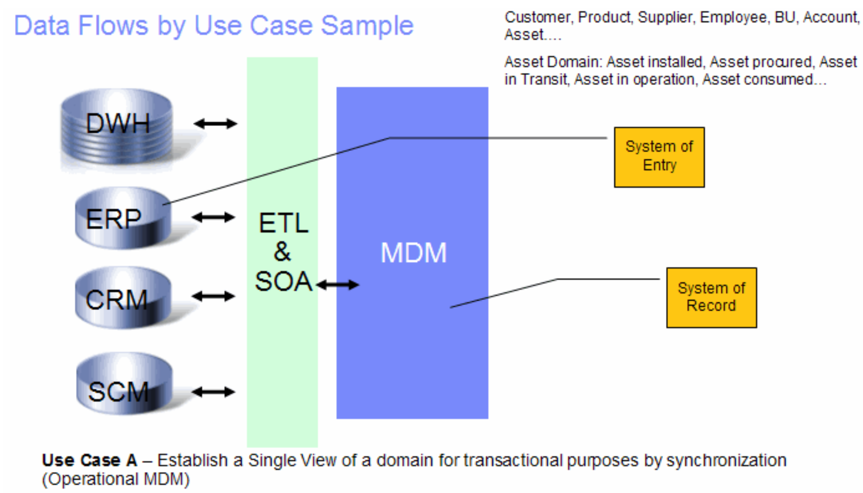

Figure 4. Illustration of use case A, aka opMDM (Single View for OLTP & BI Facts)

A more scalable, quantitative risk-adjusted, enterprise-wide approach *(Figure 5)* would be to create and store the product addition once, and test the changes and related options for their bottom-line effect, i.e. financial impact. Thus, the product manager would create a new variance in his enterprise product catalog hierarchy, change the required attributes, link a new or given general ledger account to each variance, sandbox the hierarchical (dimensional) changes into a BI tool environment, associate cost and revenue figures derived from similar products (maybe in other geographies) and re-run an existing report based upon the new structure. Once others, like the sales manager, CFO, etc., reviewed, commented and approved, this new product structure can be promoted as the new, auditable standard to be used by all BI as well as other operational systems, like the aforementioned real estate operations, billing or fulfillment, for example. However, here we still expect decision makers to render a verdict on the next-best-action by interpreting reports. We also assume a fairly significant effort (i.e. delay) in the Business Intelligence IT team to support ad hoc changes to reports and one-off custom hierarchy aggregations which may never get used again.

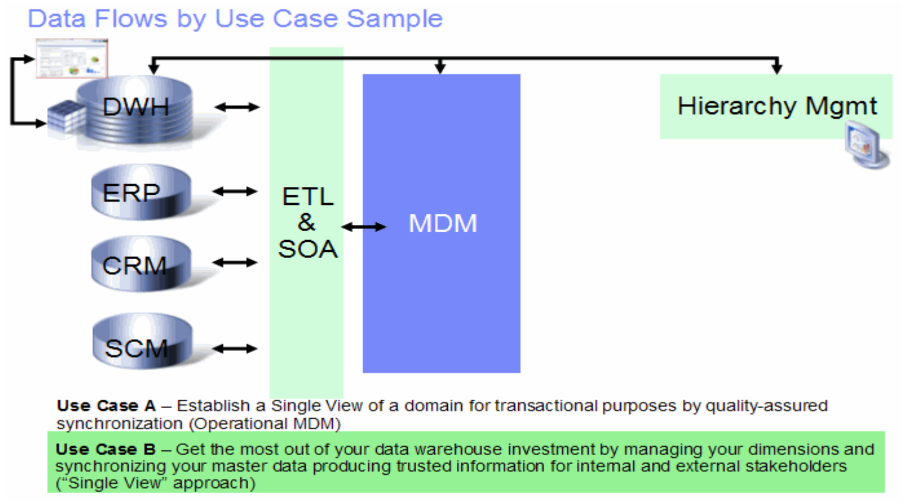

Figure 5. Illustration of use case B, aka aMDM 1.0 (Single View for Transaction & Hierarchy Management)

A Healthcare Industry Example

Similar to how law enforcement knows all relationships of a suspect and the gaming industry knows who knows who and how before a player even enters a casino/hotel, payer organizations should know a healthcare provider's corporate leadership, ownership and relationships. This information should be held and managed centrally as an important tool to combat potential repeat abuse of the payer system by provider organizations with cross holdings, hidden relationships and officers previously barred from reimbursement due to fraudulent activities emerging soon thereafter under the umbrella of an approved corporate entity. They key here would be to identify potential bad apples at the time of accreditation but at the very least at the time when suspicious transaction patterns emerge. A combination of catalog and identity resolution capabilities can minimize risk factors associated with phantom billings, etc.

COMPLIANCE

The same process, although more complex on the data structure side, would be the creation of different financial consolidation reports of a central GL instance. (White, 2011) One could be for internal management consumption, another for regulator A, yet another for regulator B or an insurer or even an investor. Many globally operating financial institutions now cover close to a hundred regulators with sometimes quarterly changing reporting requirements. Legions of developers and financial analysts now have to understand these changes, acquire and manipulate the data, prepare the reports and outfit them with meaningful footnotes so regulators' questions are minimized. Abusing the capabilities of a data warehouse may be a quick fix to fix this metadata rather than master data problem but a long time head ache in the making, particularly when the same structures are used in an operational sense. (White, 2007)

PRODUCTION AND MAINTENANCE OPERATIONS

The same concept implemented with less manual interventions and interpretation *(Figure 6)* can provide a financial view of asset-generated data in the production of a product, which needs to adhere to an SLA-compliant quality standard carrying potential financial penalties for late, incomplete or sub-standard delivery. Here we are looking at a large volume of aggregated production/usage data in combination with key reference data in the customer, product, supplier, carrier, etc. domain to derive the best course of action.

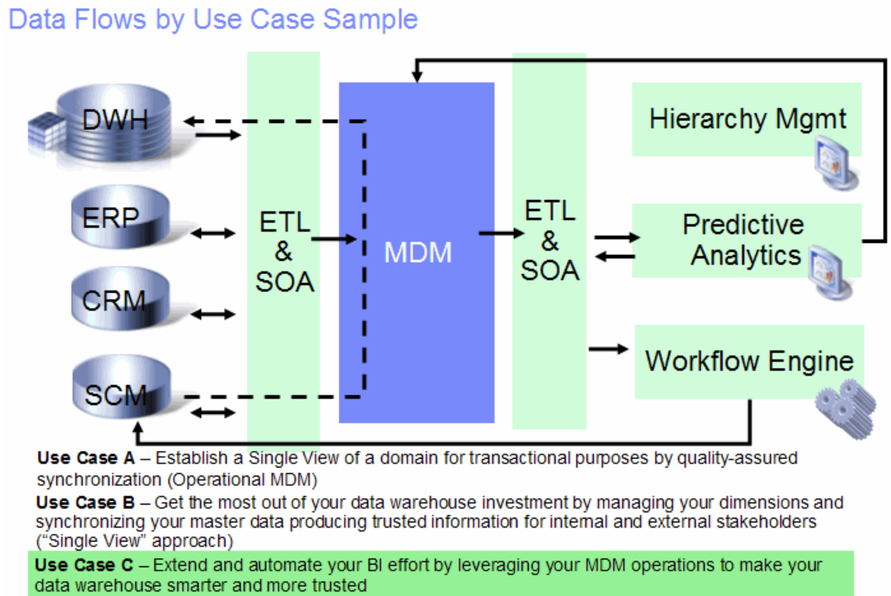

Figure 6. Illustration of use case C, aka aMDM 2.0 (Single View for OLTP & Hierarchy Management & Predictive Analytics with Process Automation)

A Mining Industry Example

A large mining company with overseas operations implemented a standard hierarchy management tool to reign in its financial reporting from various operations and increasingly more complex wholesale and retail packaging strategy. Whereas the occasional retail customer could just swing by the local plant to fill up his pickup with limestone bags of 20 kilos, very large orders for industrial and home improvement companies would be filled via trucks and trains in a wide variety of packaging and quality options. In a later step, plant operations would also treat product inventory based upon its product type, quality and stage in the manufacturing cycle using source classification codes (SCC), for example, product located, quarried, rushed, screened, stored, classified, pulverized, screened, milled, cooled, milled and packaged. *(Figure 7)* The small BI team with key stakeholders from sales and operations was charged with capturing all locations, product, asset, carrier and customer hierarchies to ensure a cohesive picture for financial and operations reporting to make future investment decisions. By including actual sensor data from machinery, automated decisions can also render additional benefits. The first effect of this multi-year, incremental project was that service level agreements with large customers, often tied together with large penalties for sub-standard quality samples or late deliveries, were effectively decreased by optimizing delivery and product schedules. In addition, carrier pickups were now scheduled more effectively without hours or days of idle trucks and trains waiting to be loaded. Moreover, a large volume of sensory data around pressure, throughput, temperature, vibration, etc. allowed production assets to be actively maintained.

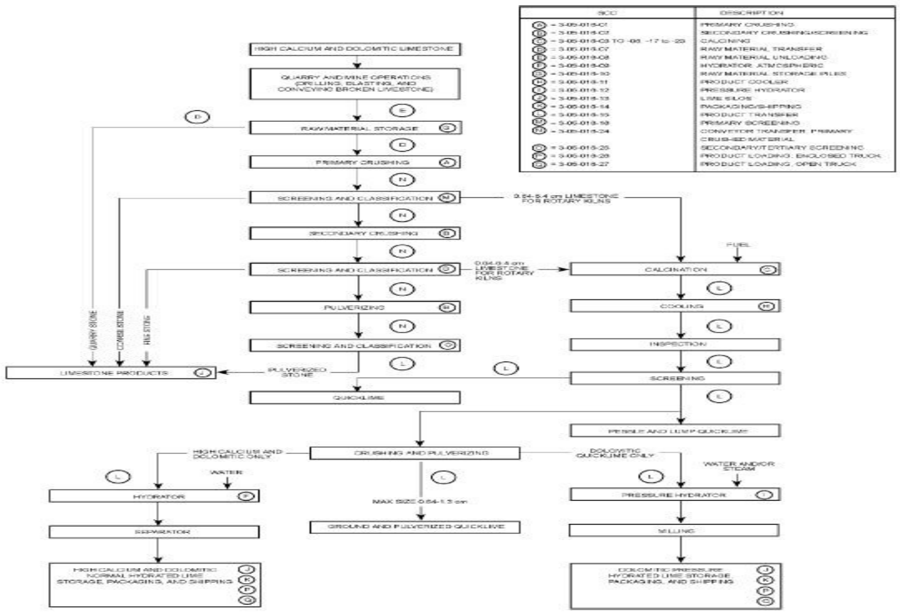

Figure 7. Limestone process flow diagram based on SCC codes – a hierarchical product structure (Source: EPA.gov)

A Discrete Manufacturing Industry Example

As discussed earlier, product attributes and their central cataloging and delivery across multiple departments and consuming applications are important depending on the number of variances and product complexity. While a producer of rare earths may have limited complexity but a lot of quality variances and locations with long contract terms, a printing, bearing or tunnel bore equipment manufacturer may only have one plant but enormously complex systems, assemblies, sub-assemblies, OEM parts with supersessions and aftermarket alternatives. More complex systems are also typically more prone to support a vivid aftersales business model requiring an even more stringent approach to the installed equipment environments and linkage of multiple catalogs to support effective onsite maintenance. As an example, industrial pump manufacturers are now slowly getting into aftersales service as an additional revenue stream. Unlike manufacturers with tight controls over who maintains their field installations, e.g. escalators, where equipment can alert the company and its partners about parts about to fail (based on sensor data), pump manufacturers find themselves in a different set of problems. They are less effective due to the fact that pumps are often installed in remote locations, have been maintained by many different providers with a wide variety of skill sets (often zero) who have installed a mix of OEM, aftermarket off-the-shelf and custom-built replacement parts making cost-effective delivery of good parts data and repair manuals critical.

In addition, missing, incomplete or false classification (tariff codes) and related documentation associated with imported products and materials can not only result in surprises in the field but also prior due to fines and inventory held up in ports for custom inspection. The ability to link complex catalogs with semi- and unstructured content in addition to predicting when a pump will fail is an obvious value proposition.

SALES & MARKETING OPTIMIZATION

In absence of growing the market through entering a new geography or acquiring a competitor or partner, the main avenues of optimizing sales and marketing operations center on how to do more with the same (or even less). This means how to gain a larger mind and wallet share with existing customers or how to incrementally take customers from the competition organically using new technologies. Both involve strategies on how to gather, segment, approach and ultimately treat clients differently based upon their long-term importance to the business.

A Retail Industry Example

A retailer may use the same approach to continuously refine a new business model, for example, if a certain group of customers who are particularly valuable to the organization due to their profitability or life time value (LTV) should receive preferential treatment, e.g. via a higher point multiplier in an existing loyalty schema, special discounts, automatic warranty upgrades, replace versus repair *(Figure 7)*, etc. Whenever a transaction pushes a customer or his/her segment's LTV by five points, he/she or the total segment may be moved into a new customer hierarchy with related benefits. This constant re-segmentation based on hierarchies can be thought of as a platinum-gold-silver-base traveler structure we are all to familiar with in the airline industry. This cycle allows marketing campaigns as well as point-of-sale personnel to always treat the customer in the most beneficial manner, for example for a replace vs. repair decision, achieving not only customer satisfaction and profitability goals but also staff encouragement. Additionally, field operatives need to be aware of what SKUs (Stock Keeping Units) are fair and cost-effective replacements for one or more catalog items as in many products, particularly technology products, life cycles move very quickly. A four-year old 50 inch HDTV flat screen with 60MHz, 1080dpi can now be replaced with a like product for 30% less making it potentially more effective to replace and sell its parts on the aftermarket then to repair it. Alternatively, the retailer could even offer a gift card delaying the expense potentially until a cheaper alternative becomes available, building an incentive for an upgrade or saving on an unspent balance.

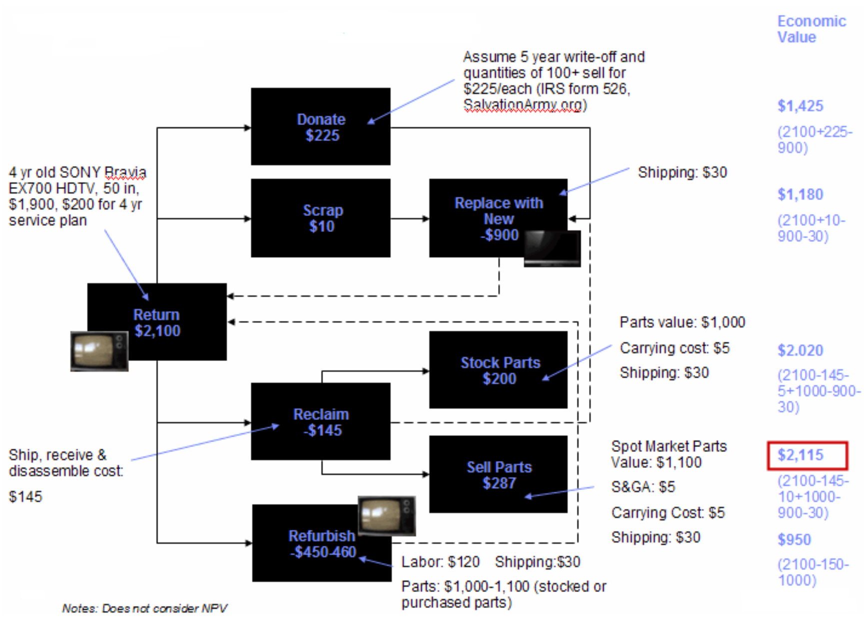

Figure 8. Replace vs Repair Logic Example using location, HR, product and customer master (& meta) data

A Telecommunication Industry Example

As suggested by research firm, Heavy Reading, it is much more cost effective to combine new insight to manage increasing network traffic and resulting customer experience challenges than to invest in more bandwidth. Thus, they suggest to telecommunication providers to improve their average annual 6-9% revenue leakage and combat subscriber churn challenge by incorporating periodically aggregated usage data (Call Detail Records), probe data, maybe even social media intelligence and combine it with key customer KPIs (and Key Performance Predictors, aka KPPs). (Banerjee, 2011) Custom business logic can pull data in stream and batch fashion from low latency operational systems and move it to a master data instance via the enterprise service bus for analysis of these aggregates to suggest the best course of action when engaging a subscriber segment. In industries with little or no brand loyalty which need to seize the "moment of truth" with a client, this could and should occur in real-time at the store, online and call center touch point. *(Figures 9 and 10)* It can also occur automatically at the cell site level when a preferred subscriber segment places a call or data request on a new or high-profitability service, like IPTV.

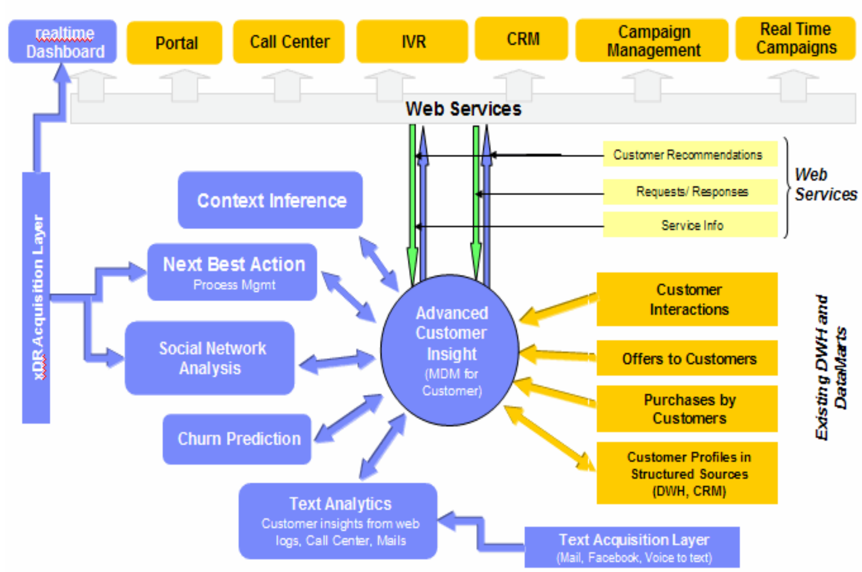

Figure 9. Illustration of solution topology - MDM centric sample of a customer analytic hub accessing transactional data for analytical interpretation to serve to consuming customer touch-point applications (source: IBM)

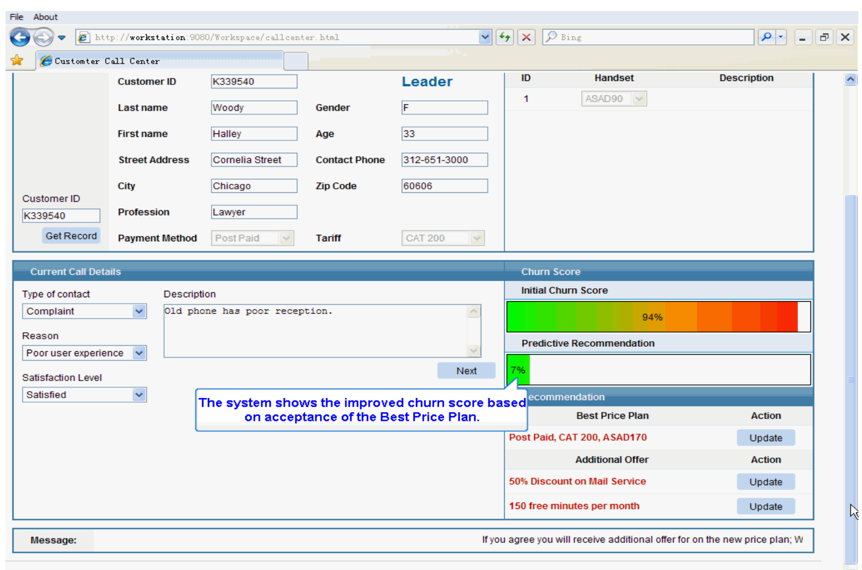

Figure 10. Call center application user interface working a telecommunication customer churn issue via traditional customer and analytical MDM data utilizing predictive logic (source: IBM)

A telecommunication operator may also choose to track improvements in its understanding about a household's individual users (a hierarchical structure) of the service and relate it to the improvements in churn or product upgrades purchased under a single account (the "Single View of Customer", aka "Customer 360"). Sometimes the household can also be represented by a set of loyalty program account numbers which unlike with retailer's single number should identify an actual set of "users" (i.e. the children) instead of just the "bill payer" (i.e. parents). In such a scenario, previously unknown values for each user will improve marketing campaigns, store, web and call center interactions by being able to associate actual phone numbers with newly available attributes of the true users of the service, not just the payer of the invoice. Examples could include their call and download preferences (time of day, duration, location, genre), association with the actual bundle or bundle components (aka product) being used and correlating it to like peer groups.

SUPPLY CHAIN OPTIMIZATION

Similarly, organizations interested in improving their time-to-market when testing or introducing a new product variance (location, bill-of material changes, etc.) may enhance their analysis capabilities native to a product catalog platform. Dependent on the strength of the organization or the level of integration in the sector, additional data aggregates like supplier raw materials, work-in-progress (WIP), in-transit, in customs, etc. data can be added to shorten the introduction cycle and optimize investment allocation. (*Figure 11*) Ultimately, the commercial goal of an end-to-end supply chain visibility from warehouse demand, distribution center replenishment, raw materials storage, third party factory floor to in-transit operations is to decrease inventory stock piles at every stage to achieve just-in-time inventory (JIT) to allow for available-to-ship (ATS) and available-to-promise (ATP) processes. Additionally, marketing departments benefit from it by starting promotions before inventory actually arrives at its factory doors.

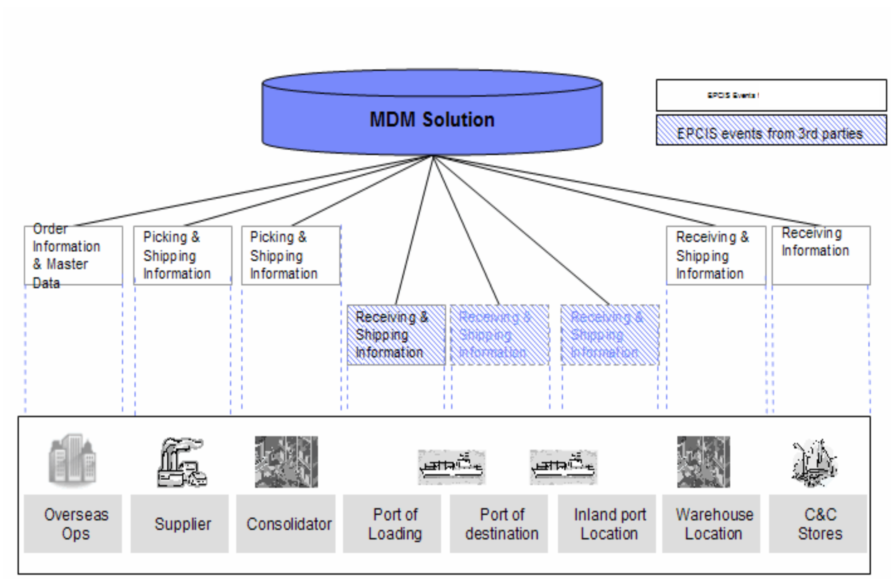

Figure 11. Global supply-chain depiction fed by an MDM instance (source: IBM)

ASSET, EQUIPMENT & PRODUCTION OPTIMIZATION

The decision when, where and how to physically and financially deploy assets, which can range from ten thousand to a billion dollars these days, has a major impact on organizations. Often, it is close to a do-or-die judgment call when variables are numerous making risk assessments an iterative set of complex calculations when new data becomes available.

An Oil & Gas Industry Example

Asset-centric operations, such as a petrochemical company, may use the same concepts in collecting, structuring and provisioning a master repository with historical data from geological and crude samples, drill sensors, well activities such as maintenance and production data, well logs and profiles and correlate them with contractual obligations and safety regulations. Hierarchical changes to the asset due to its life cycle need to be reflected in reporting to decrease operational expenditure and provide the true equipment situation in the field to ensure employee safety. However, it will also assist in decisions on changes to production throughput, preventative maintenance, labor allocation or drill vectors. The relevant data and decisions depend on the information consumer, which views the same asset differently by function, i.e. IT, drill vs process vs reservoir engineer, geologist, and financial analyst.

Similar to the aerospace sector, rigs are typically one-of-a-kind combinations of "off-the-shelve" and highly customized assemblies, i.e. product hierarchies. Just as every Boeing 737-800 delivered to ten different airlines has its individual toilet configuration, lighting setup, etc. every rig's construction is highly custom due to where and how it will spend its next twenty years producing hydrocarbons. Just as every airplane comes with its 500 kilos of manuals (now in CD format), it is important that unstructured information is increasingly structured and electronically updated using a variety of XML-based standards, e.g. SGML, to assemble a relevant, short instruction for a particular repair job with visual part descriptions on the fly.

Within an upstream scenario, for example, a reservoir engineer may be interested in 75% of the same data as a geologist ultimately trying to render an opinion on reservoir distribution, size, salinity, porosity and ease of access to pump the typical 75% of available, yet with today's means technically unreachable hydrocarbons, to extend the life of a well. In the past, he accomplished this via running many simulations on key data from data historian systems but nowadays he has the ability to also tap real-time well-sensory and production data from 30,000-90,000 sensors on a rig. Such analysis has an impact on financial reporting and capital investment decisions on where to drill the next well.

On the other hand, a drill engineer may be more concerned about drilling and completion and as such a well's profile, maintenance and production history leveraging historical and real-time subterranean sensory data to assess when a particular piece of equipment is about to fail to ensure continued production or safety for the rig crew. (*Table 1*) In the past only about two dozen of these 1,000 sensors were used to feed databases and these would be purged every two weeks due to capacity. This aspect is important as his environmental and safety colleagues need to be able to produce longer auditable safety statistics to regulators. Also, a failure to produce for a day due to a long mean time to repair (MTTR) for a malfunctioning valve or pump (note: many firms outsourced parts warehousing to third parties) can cost a producer millions of dollars. Similarly in a reservoir-centric perspective, failure to locate the next bore site in time will cost producers close to a million dollars a day in contractual obligations to rig operators.

Point of View (Role)	Well	Well Origin	Wellbore	Wellbore Completion	Wellbore Contact Interval	Production Stream
Geoscientist	X	X	X		X	
Drilling Engineer	X	X	X	X		
Petrophysicist					X	
Production/Completions Engineer				X	X	
Reservoir Engineer				X	X	X
Facilities Engineer	X	X				X
Field Operator	X	X				X
Production Accountant	X					X
Land Management	X	X				X
Financial Analyst						X
Data Steward	X	X	X	X	X	X

Table 1. Data domain interest per function (source: IBM)

Unlike in a highly non-standards oriented sector like retail or highly standardized industries like banking and telecommunications, the petrochemical sector has a multitude of competing standards (CIM, PPDM, ISO) making the push for a central way to manage a semantic and data model crucial. In this instance MDM can facilitate this process and enable its implementation.

A Telecommunication Industry Example

Similarly to the previous oil & gas scenario, it is important for all divisions of a communications company to make the right investment decisions when it comes to subscribers, partners, content and network equipment. Only if key transactional (aka usage) data about subscribers is captured and analytical insight (historical and predictive) about them is aggregated to drive real-time decisions, an operator can assess how to prioritize repair crew dispatch, bandwidth allocation and potential regulatory exposure due to an outage.

Since today most new business models include location based services it is crucial that every affected server's utility is maximized serving high-margin, high-revenue content to the most profitable subscribers, by giving priority to contracts for high-impact content served by key partners. Such third parties may receive subscriber information from the operator for accurate targeting and/or paying it for use of their charge and billing management as well as managing software versioning on its devices. In addition, if a central view of all services, products, offers, charges and resources exist, legacy service and product cannibalization by its latest version can be assessed across all channels.

A Healthcare Industry Example

On the healthcare side, such enrichment with analytical data would encompass a subset of Agency for Healthcare Research and Quality (AHRQ) measures, such as post-op infection rates, chronic disease progression, return admissions, etc. A change in this measures, often published by community hospitals to attract local government and investor support for future capital investment, will be increasingly mandated to also warrant national health care system reimbursement, e.g. Pay for Performance schemes for Medicare in the US. However, the same data can find its use in clinical trial study design, patient recruitment, cohort identification, exploratory analysis, patient safety, physician pay schemes, patient flow and capacity management. Similarly, provider cost in conjunction with payer administrative cost turn into a medical loss ratio which constitutes the percentage of premiums spent on actual care rendered. Current ratios in the US are all across the board and stir up discussions around a mandated minimum of 85% (Familiesusa.org, 2008; Healthcare.gov, 2011), privatization versus nationalization of key healthcare programs. Increasingly, local, state and national governments will set this ratio to ensure administrative expenses are decreased via process and technology efficiencies to ensure the vast majority of premiums are spent on treating the patient rather than satisfying the investors.

WHY THE STATUS QUO WON'T WORK

The status-quo assumes that hierarchies are to be generated in a (semi-)static fashion for the sole purpose of being populated into a warehouse environment for reporting, ill-equipped for the real-time and federated decision-making demands of today's business models with many individualized transactions and many users operating in rapidly changing business models. In reality life is more complex. In a typical scenario a project manager requests a freebie from his manager for a particular customer. The first question he will encounter is, "who is this customer to us?" followed by "how much revenue is he generating, how important is he as a reference, how much is he costing us and how much will he bring us in the future?" All these questions should be answered by Key Performance Indicators. Often, this decision should be automated, especially if the underlying data represents thousands of transactions and the decision has low risk but potentially high impact.

Even more so, the way the decision driving data is generated varies based on when, who and how different haphazardly-maintained data silos are accessed to create reports for the same audience. These silos employ a small army of developers for weeks every quarter to assemble regulatory reports based on quarterly changing requirements. Today, all KPIs/KPPs are manually viewed in a hard copy fashion or browser dashboards for manual decision making, despite that by the time the report is created it is already old news and a competitor has already reacted to a trend he has (un)intentionally anticipated. The to-be anticipated financial impact of changes to customer, organization or product hierarchies, say based on an acquisition, supplier change or restructuring, cannot be optimized. Transactional (trading, settlement, billing, service consumption, point-of-sale) and predictive applications are in dire need of hierarchical information and its ever evolving attribution, say a new attribute sourced from a BI application which would improve customer segmentation.

These shortcomings ignore the interlock required in an end-to-end process, the fact that some decisions should be made automatically based on KPIs/KPPs changing sometimes by the hour and that a single hierarchy instance can and should serve many variations based on the use case dictated by the type of system or user expectation.

IT'S A JOURNEY – STAY ON TRACK

Similarly to how Henry Ford established a new frontier in engineering by combining the old with new methods and thinking and the snowplow multiplied the impact of the automobile, we need to promote smarter data use by utilizing what we have invested in already. If you do not have or envision having and maintaining a top-notch operational master data management application, work with clients to specifically address a set of key issues they need to overcome step-by-step to make incremental inroads on putting their legacy data to work as master data. We can leverage some key features in available COTS (commercially off-the-shelve) software products to stage, cleanse, consolidate, move, transform, enhance, structure and publish this key data family to empower downstream applications and their users. I suggest to initially focus on a single use case involving a limited number of products in a single location to empower a key customer process to warrant better financial results. (*Figure 12*) The point is to avoid an ERP-style top-down mammoth project. The problem should be worked backwards from the user point-of-view limiting the initial effort based on a single process or application, which can then be grown over time by adding additional use applications, processes and data domains.

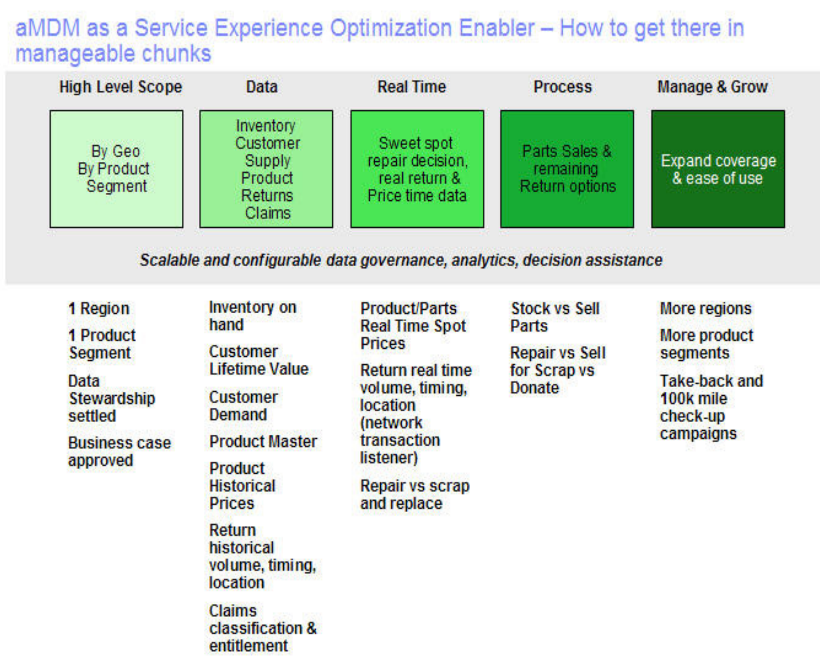

Figure 12. A proposed approach on how to incrementally grow an aMDM 2.0 strategy for a retailer

More importantly, we should not forget that an ongoing investment in data governance involving people, process and technology as key ingredients is a must-have. An initial investment into only one of these or with a set or unintended expiration date will not only deteriorate initial benefits over time but increasingly bad user experiences will make it even harder the next time around to harness the power of your data, which are just another commercial asset of the business. Customers need to understand that once this journey starts, there is no going back to the old ways as use cases emerge and transform the understanding of which additional technological components are required, which governance processes need to be instituted and who should have a say in them. (White, 2011)

REFERENCES

Banerjee, Ari (2011) “Plugging the revenue leaks”, Heavy Reading, 2011.

Byrne, Brian (2010), “Data warehouse centric topology of Analytical MDM”, IBM, 2010.

English, Bob (2009) “I will build a car for the great multitude”, The Globe and Mail, March 30, 2009.

Familiesusa.org (2008) “Health Policy Memo: Medical Loss Ratios: Evidence from States”, June 2008.

Hayler, Andrew (2010) “MDM Market Review - Interview with IBM MDM Product Management”, Informationdifference, 2010.

Healthcare.gov (2011) “Medical Loss Ratio: Getting Your Money's Worth on Health Insurance”, November 22, 2010.

Hechler, Eberhard (2011) “Integrated MDM-DW Architecture Platform“, IBM, March 10, 2011.

White, Andrew (2009) “The Differences, and Similarities, between Operational MDM and Analytical MDM”, Gartner, November 10, 2009.

White, Andrew (2011) “New “use case” for MDM emerges with “middle-stream” MDM?”, Gartner, January 21, 2011.

White, Andrew (2011) “When MDM isn’t MDM? In Finance of course, well sometimes…”, Gartner, March 1, 2011.

White, Colin (2007) “Using Master Data in Business Intelligence”, BI Research, March 2007.

IBM Software Group

Route 100

Somers, NY 10589

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.