Bloor

White Paper

White Paper by Bloor
Authors **Philip Howard**
Publish date **June 2017**

Single Customer View

> **Without going into GDPR in detail, companies affected have a choice: they can either implement siloed solutions multiple times across their organisation, or they can build an SCV and then implement a single GDPR solution against that SCV. Given the timescales involved, implementing a hub-based MDM solution will simply not be practical, even if it was deemed to be cost-effective.**

Introduction

This paper is about constructing a single customer view (SCV). That is, a single, consistent perspective of the data pertaining to individual customers (or clients, patients, stakeholders et al), across your entire business. This is hardly a new concept and the implementation of such systems has been largely focused on master data management (MDM) systems – at least in large organisations – over the last few years. However, MDM has always been complex, costly and time consuming to implement. As a result, there have always been companies interested in building SCVs without any MDM implementation. In our view, this interest is accelerating and it is driven, not only by the cost advantages of eschewing a formal MDM system, but also by two other forces. The first of these is regulatory and the second is technological. Considering the latter first, there is a move away from traditional data warehousing (towards data lakes), which is often where MDM is implemented. More importantly, the advent of regulations such as the EU's GDPR (general data protection regulation), which comes into force in May 2018, is driving the adoption of SCVs. Without going into GDPR in detail, companies affected have a choice: they can either implement siloed solutions multiple times across their organisation, or they can build an SCV and then implement a single GDPR solution against that SCV. Given the timescales involved, implementing a hub-based MDM solution will simply not be practical, even if it was deemed to be cost-effective.

Having made the preceding statements, we do not want to dismiss the value of a formal MDM installation. What we intend to do in this paper is to discuss and contrast what can be done to build an SCV using data quality and associated tools as opposed to using MDM products. That said, we will also consider the relationship between these two approaches, because data quality approaches should really be thought of as synergistic, and possibly a first step, to MDM. They are complementary, but also differ in many ways. In particular, data quality has a role to play that is distinct from that of master data and which may be required either as a stand-alone function or as a precursor to implementing master data management. This has important implications for the selection of products in these two categories.

In this paper, we will consider the part that these two technologies play in information management in general, the ways in which they complement one another and when you should consider one (data quality) in preference to the other (master data management), with particular reference to building an SCV.

A single customer view (SCV) is a single, consistent perspective of the data pertaining to individual customers (or clients, patients, stakeholders et al), across your entire business.

Data quality

> One European bank was able to save €50m on its capital adequacy requirements under the Basel II regulations when it was able to prove the reliability of its data.

Data quality products consist of two parts: data profiling and data cleansing. We need to discuss each of these separately.

Data cleansing

Despite its name, data cleansing actually does three things: it ensures accurate data (cleansing), removes duplicated data (matching) and supplements data (enrichment). Taking these in reverse order, enrichment helps to make data more complete. For example, you might go to a D&B website and retrieve a SIC code to add to a customer record, or you might retrieve latitude and longitude information against a postcode, or you might get credit information from Experian.

Matching is used to remove duplicate records while cleansing helps to ensure that the data in the resulting record is accurate. There are a number of reasons why either or both of these functions might be important to your business:

- **Good data quality can save you money.** For example, one European bank was able to save €50m on its capital adequacy requirements under the Basel II regulations when it was able to prove the reliability of its data. Reduced mailing costs from the removal of duplicated data, is another such benefit. Conversely, lack of good data quality can cost, both in financial and other ways. For example, Gartner estimates that bad data can cost large organisations $14.2 million per annum. In a recent survey conducted on behalf of Experian, the company reported that *"Almost three quarters of our survey respondents (72%) agree that data quality issues impact consumer trust and perception, with 64% reporting that inaccurate data is currently undermining their ability to provide an excellent customer experience."*
- **Good data quality can make you money.** For example, Sallie Mae has estimated that introducing an internal data quality project enabled an increase in revenue (through better targeting) of over $2m based on increased loan volumes of around $50m. At the same time the company saved between $4m and $5m by switching from postal to email-based marketing, which was similarly enabled by its use of data quality.
- **Good quality data may be required for compliance reasons.** For example, the EU's Solvency II regulation for the insurance sector requires that data be *"accurate, complete and appropriate"*. The MiFID II regulation for capital markets uses the same terminology. The Financial Services Compensation Scheme in the UK also requires the provision of accurate information from deposit holders. A completely different example is the British Army. It discovered, as part of its data quality exercise, that it had a foreign national in the army who was not legally allowed to be a member of the British Army. The discovery of this fact, through the use of data quality techniques, meant that the issue could be resolved without causing a diplomatic incident.
- **Good data quality enables a variety of IT processes**, including data migration; loading data into a data warehouse, mart or lake; and data archival. Without good quality data, these processes are liable to fail. For example, in surveys conducted by Bloor Research into data migration, the most frequently cited reason for overrunning projects was the combination of either *"poor data quality"* or *"lack of visibility into data quality issues."*
- **Good data quality similarly supports a variety of business processes.** For example, the British Army used to have issues with its personnel records, which meant that privates – who may be troopers, artillerymen, drivers, guardsmen and so on rather than privates per se, depending on their regiment and skillset – could easily be wrongly assigned to tasks, both in the barracks and on the battlefield. More generally, if you don't have a common nomenclature then processes can get delayed, backlogged, bottlenecked or fail entirely.

- **Good data quality makes it possible to accurately report on, and analyse, information.** For instance, Emerson Power has manufacturing plants all around the world which historically ran independently. When the company started to implement data quality processes, it discovered that it had multiple plants manufacturing the same product even though these had different product codes and different descriptions (in different languages). Using data quality technology, it was able to match these products and, for the first time, actually quantify the sales of this one product. As corollaries:
  - Once you have this information you can start to consider manufacturing plant consolidation. Multi-sourcing may be fine but seven or more plants making the same product (which was the case with some Emerson products) may be too many. A similar argument would apply to supply chain consolidation.
  - Emerson reported a significant reduction in staff turnover after cleaning up their data. This is logical: if you have to make business decisions (which you will be blamed for if they go wrong) based on data that you don't trust then that causes stress. Stress results in low morale and low morale leads to high staff turnover.
  - A related requirement is to be able to consolidate across multiple bills-of-materials. Procurement can miss contract opportunities if they don't fully appreciate that they are ordering the same part over and over again.

While we are getting ahead of ourselves it should be noted that all of the preceding discussions are based solely on data quality processes and have nothing to do with master data management per se.

Data profiling

Data profiling tools are also used for a variety of purposes: to discover errors in your data, in which case it acts as a precursor to data matching and cleansing; to monitor the data for errors on an ongoing basis to support data governance; to discover patterns of data (for example credit card numbers) that need to be masked or anonymised for data protection purposes; and to support the discovery of relationships (which may be implicit and not defined in, say, a database schema) in the data, both within a data source and across data sources.

In the case of relationship discovery this is important for data migration and archival purposes because it is essential that these relationships are maintained when you move the data, and it is also important, as we shall see, in supporting master data management implementations. Further, in very large environments it is a necessary pre-cursor to consolidation and rationalisation. For example, AT&T has tens of thousands of databases and it has recently been using data profiling technology to discover data relationships across all of these sources. While AT&T has not undertaken this process specifically to support the creation of an SCV, such a process of cross-source discovery is necessary in order to enable the creation of an SCV. It is worth adding that this same process will be needed in order to support GDPR. regardless of whether you want to use the SCV for commercial benefit, it will be necessary to support access, change and *"forget me"* requests from customers.

It is worth adding that this same process will be needed in order to support GDPR. regardless of whether you want to use the SCV for commercial benefit, it will be necessary to support access, change and *"forget me"* requests from customers.

Master data management

Master data is data that needs to be shared by multiple systems or business processes. For example, a third-party company may be a customer of ours, but it actually may also supply another part of our own organisation, and so also be a *"vendor"*. A product line that our company sells may have a unique code which is stored in a marketing system, but it also used by systems in manufacturing, sales and logistics. Examples of master data include customer, product, asset, location, employee, organisational unit, legal entity and chart of accounts. Data that is not shared is not master data: for example, a record of a specific sales transaction is important, but in itself it typically will not be shared with other systems.

Critically, transaction data is stable: you go into a store and buy a bar of chocolate at a certain time and day for a given price. That transaction happened and does not change, but the context of the transaction may. In a few weeks that store may be switched into a new sales region, the chocolate bar may be re-classified by marketing into a new *"luxury foodstuffs"* hierarchy, the store itself may even be closed or sold on, but that transaction still happened on that day at that time. This is a vital distinction, because it is the way that master data changes that causes business problems.

So, master data is that data that defines who a customer or supplier is (name and address, contact details and so forth), what a product is (code, description, weight, volume and so on), what contracts consist of, details of plant and machinery, company location information, and, generally, defining details of any business entity that is relevant to the company. The storage, processing and administration of this master data is typically the role of master data management (MDM) or specific subsets of MDM such as customer data integration (CDI), product information management (PIM) and global supplier management (GSM).

MDM is, of course, the management of master data. The need for it comes particularly from disparate, overlapping systems that are inconsistent with one another. This may occur because of siloed developments, as the result of mergers and acquisitions, because you have some applications hosted in the Cloud and some in-house or simply because of a proliferation of systems. However it derives, MDM is about ensuring the consistency of master data across these multiple applications. This applies in technical terms in ensuring that if, for example, you have six different CRM systems that all deal, or potentially deal, with the same customers then you want those records to be consistent.

Consistency also applies at the business level: if a customer of your bank has a mortgage with you then he or she will no doubt also feature in your marketing database. The record in that database should be flagged not to send the customer marketing literature about a potential mortgage. Otherwise it will cost the bank money and annoy the customer. So, the business needs to be consistent about its use of data across multiple applications and needs to have some method (MDM) of integrating data across those applications. A third way in which MDM relates to consistency is when it is used to store reference data that ensures that applications share a common (and therefore consistent) definition of relevant reference data.

A further function of MDM is to provide hierarchy and relationship management. A customer, if it is a corporation, may have multiple subsidiaries, each with multiple addresses and with multiple touch points (people) within each of those locations. Thus, from a high-level perspective, there may only be a single customer but from an operational point of view there may be thousands of individuals who each need to be treated as representative of this customer. This secondary function of MDM is therefore to maintain accurate and consistent information about this hierarchy within the context of the customer as a whole. Of course, similar principles apply with respect to product and other types of hierarchy that may be managed via MDM implementations.

The actual architecture of MDM implementations varies but, typically, either relevant data is stored once and shared by participating applications, or the data remains with its original application but pointers are used to update other data sources that need to share the same information when any of the participating data sources is updated. The first of these approaches is generally known as a hub while the latter is called a registry style system. There are also hybrid approaches whereby mastering such as the coexistence model whereby mastering remains at the source system level but a master copy is held centrally for synchronisation purposes.

Complementary technologies

There is little point in ensuring consistent data across multiple data sources if that data is inaccurate. One of the biggest challenges in call centres, to take an example, is first time resolution. This is the idea that the operative that takes the call should be able to resolve the caller's issue as often as possible. This is because call escalation (passing the call onto somebody else or requiring a second call) significantly increases the cost involved in handling calls. To maximise first time resolution the call centre operative needs complete and accurate information about you as a customer. If there are multiple data sources involved (because, say, you have a checking account, a mortgage and insurance policies with the bank) then MDM will be required to bring all this information together but that will be of little value if the information is incorrect.

Data quality is also important with regard to hierarchy management such as bills-of materials, organisational hierarchies and so on. The operational views that hierarchies provide must be consistent, complete and accurate: each individual entity within the hierarchy needs to be accurate and its position in the hierarchy (its relationship with other entities) has to be correct, otherwise the high-level view of corporate or product structure will be inaccurate.

Regardless of whether hierarchy management is relevant (and that is not the case for all companies), in practice we do not know of a single user of MDM that has not simultaneously or previously implemented data quality processes to support their master data.

Further, data quality is not only complementary to MDM in terms of data cleansing but also de-duplication. This applies specifically to hub-based approaches to MDM where a single instance of the master data is created. Clearly, if you are creating a master data hub for say, CRM data, then the same customer or prospect will appear multiple times in those different systems and you will need to merge these into a single record, which will require the matching capabilities of an appropriate data quality tool.

It is not only data quality that is complementary to MDM: the same applies to data profiling. There are both general and specific reasons for this. In the general case, you can treat MDM hub-based implementations as a form of data migration, and data profiling has an important role to play in easing the course of such projects. Further, when it comes to specifics there are a number of pre-cursors to MDM that can most easily be achieved through the use of profiling. For example, you need to determine matching keys across your data sources and you will also need to do precedence analysis (that is, deciding which sources you trust most for particular pieces of information). Data profiling can help to automate these and other processes involved in the implementation of MDM.

We would argue that data quality (including data profiling) is absolutely necessary for the successful implementation of MDM. However, the reverse is not true.

“Data quality (including data profiling) is absolutely necessary for the successful implementation of MDM. However, the reverse is not true.”

Data quality is distinct

Both data cleansing and data profiling are used for a variety of purposes that do not involve MDM. In the case of the former we have already mentioned its use in supporting data migration, archival and other data integration tasks as well as its role in monitoring on-going data quality within a data governance context. In so far as data cleansing is concerned there are several points to make:

1. **Consistency is only important when data is duplicated across source systems.** If you are concerned about the accuracy of your personnel records, as in the case of the British Army cited above, then master data is irrelevant because if you have two different personnel systems then they would have different people in each one and therefore there would be no issues with consistency.
2. **Data cleansing can be used to prevent poor quality data entering your system.** For example, you can implement real-time data quality alongside data entry processes for inputting data into, say, an Oracle eBusiness Suite application, checking that the data is entered and is correct. This is very important from a *"prevention is better than cure"* perspective.
3. **In some cases, MDM is either not necessary or may be useful only at a later stage.** If we take the case of Emerson Power, as referred to previously, the key issues were sales reporting and the potential consolidation of manufacturing facilities. These can (and were) achieved through data quality cleansing and matching processes and the implementation of standardisation for product codes and descriptions. Certainly, this standardisation could be achieved through the later implementation of MDM but it is not strictly necessary, especially if appropriate data governance procedures and policies are in place whereby you monitor that policies around the standardisation of product codes and descriptions are adhered to.

So, it is evident that there are use cases for data quality that do not rely on master data. Moreover, it should be self-evident that where the converse is the case, then high quality data is a necessary first condition for supporting MDM implementations: there simply is no point in MDM if the master data that you are managing is not fit-for-purpose. Yes, you may be able to share customer, product or other data across the enterprise on a consistent basis but if that consistency is based on invalid information then the benefit to the business will be zero. Indeed, it may even be damaging if people assume that the information is correct when it is not.

The question that might arise, however, is the time relationship that exists between data quality and MDM: do you implement your MDM system first and then worry about quality, or is it a simultaneous process, or should you worry about data quality first?

The key question is: what is the fastest route to value? We would suggest that it is in focusing on data quality in the first instance. If you implement MDM first you won't get any value until you have also ensured your data quality, as we have already discussed; and the same applies if you attempt a parallel implementation. On the other hand, there are direct business benefits that derive from having reliable as opposed to error-prone information, so if you start with data quality the time to value on your investment will be reduced. Moreover, we could also argue that the greatest value in such a project actually accrues from the higher quality of data provided rather than through the on-going management and provision of master data. Indeed, as much as 80% of the business value could accrue from only 20% of the cost, effort and time of a '*full*' MDM program: certainly this will be true in the short to medium term and may also be true in the long term as well.

There is another reason why it makes sense to worry about data quality first and MDM second. This is that actually having top quality data may change your MDM priorities. If you are considering MDM and data quality holistically then

"What is the fastest route to value? We would suggest that it is in focusing on data quality in the first instance."

one set of solutions may appeal to you whereas if you regard data mastering and MDM separately then the reasons for adopting a particular approach to MDM (for example, a lightweight registry-style approach as opposed to a more complex and time-consuming MDM hub) may change: the cost-benefit equation may be altered.

To be more specific, we are suggesting that it makes sense to consider the value of data quality as something that is independent from MDM and, conversely, that the benefits of MDM should be evaluated excluding data quality. If you do this you may find that your priorities are altered: you may decide that PIM is more important than CDI, for example. Indeed, you may even decide to defer, alter or cancel broad MDM plans. If the data quality solution chosen has facilities to support data governance in general and data stewards specifically, then it could well be that this is all that you need.

It makes sense to consider the value of data quality as something that is independent from MDM and, conversely, that the benefits of MDM should be evaluated excluding data quality.

Considering the single customer view

> **A major driver behind the current interest in SCVs is GDPR. That comes into effect in less than a year. In our experience, hub-based MDM solutions typically take at least a year, and often longer, to implement. They are not a choice in the context of GDPR.**

What are the elements required to establish an SCV? Firstly, you need to discover all the data sources in which relevant data is located. This requires the relationship discovery capabilities of an appropriate data profiling tool. Note that we have been careful to use the word *"appropriate"* here, because not all data profiling products are very good at cross-source relationship discovery, especially where those sources are heterogeneous, and which is exactly the sort of capability that GDPR, for example, will require.

Secondly, you need to establish the accuracy of the data in each of those sources and you need to remove duplicate records. This is the role of data cleansing and matching. You may also want to enrich data sources with other data that is available either internally (for example, from your call centre) or externally (for instance, demographic or location-based data).

Finally, you need to ensure that the data in these sources, having been cleansed and de-duplicated, remains consistent and that data errors do not creep back in. The latter can be accomplished through data profiling tools but maintaining consistency is more of an issue. If you have a hub-based MDM implementation then all the data is gathered together in one place so this is not a problem. If you have a registry-based MDM system then changes in one place automatically get propagated to other relevant source systems. The most common alternative to using MDM is to rely on CRM. If there is a single, centralised CRM system then this will effectively act like a hub-based MDM environment specifically for customer data. However, there may be an issue if you have multiple CRM systems that are not linked together in some way. If you are (you should be) using data profiling to monitor data quality then you should be able to detect when related data becomes out of sync, and it then becomes the task of the data steward to update other data sources appropriately. For reference, it is estimated that data becomes stale because of new phone numbers, changed addresses, people dying, and so forth, at a rate of between 1 and 1.5% per month. If you have a million customers that's 500 changes per day, which sounds like a good argument for automating the update process. On the other hand, if you have a hundred thousand customers then 50 changes per day would seem quite manageable on a manual basis. Also bear in mind that multiple CRM systems may have only limited (if any) overlap in terms of common customers.

As we mentioned previously, a major driver behind the current interest in SCVs is GDPR. That comes into effect in less than a year and companies are expected to be compliant from day one. In our experience, hub-based MDM solutions typically take at least a year, and often longer, to implement. They are not a choice in the context of GDPR. For most companies, the choice will come down to relying on a CRM system or systems (it is arguably a good time to consolidate on a single CRM system) or deploying a registry-style MDM solution, in both cases in conjunction with an appropriate data quality and profiling solution.

Conclusion

From a generic point of view, data quality is separate from master data management. While the latter requires the former the reverse is not true. Moreover, even when implementing MDM and data quality together it is likely that the data quality tools that you license will be used for other projects and to support other capabilities across the enterprise that are not related to the MDM implementation. This, as we stated at the outset, has important implications for the selection of data quality tools as opposed to MDM products. Indeed, it strongly suggests that you should consider the choice of data quality provision separately from any MDM product selection decisions. Of course, the former will need to be able to work co-operatively with the latter but the broader remit of data quality means that even if you are only thinking about MDM right now, you should bear in mind how data quality will be re-deployed for other purposes across the organisation in the future, and make licensing decisions that are based on that premise.

As far as building an SCV is concerned, many companies that might consider MDM as a potential solution, are going to be limited in their short-term choice, because of GDPR, either to being CRM-based or using a registry. But in either case it is going to be data quality that precedes that implementation and, especially, it is going to be the discovery of the elements that make up the SCV this is going to be most important: and that is the role of data profiling.

You should consider the choice of data quality provision separately from any MDM product selection decisions.

FURTHER INFORMATION
Further information about this subject is available from
www.bloorresearch.com/update/2338

About the author

PHILIP HOWARD
Research Director / Information Management

Philip started in the computer industry way back in 1973 and has variously worked as a systems analyst, programmer and salesperson, as well as in marketing and product management, for a variety of companies including GEC Marconi, GPT, Philips Data Systems, Raytheon and NCR.

After a quarter of a century of not being his own boss Philip set up his own company in 1992 and his first client was Bloor Research (then ButlerBloor), with Philip working for the company as an associate analyst. His relationship with Bloor Research has continued since that time and he is now Research Director, focused on Information Management.

Information management includes anything that refers to the management, movement, governance and storage of data, as well as access to and analysis of that data. It involves diverse technologies that include (but are not limited to) databases and data warehousing, data integration, data quality, master data management, data governance, data migration, metadata management, and data preparation and analytics.

In addition to the numerous reports Philip has written on behalf of Bloor Research, Philip also contributes regularly to ***IT-Director.com*** and ***IT-Analysis.com*** and was previously editor of both ***Application Development News*** and ***Operating System News*** on behalf of Cambridge Market Intelligence (CMI). He has also contributed to various magazines and written a number of reports published by companies such as CMI and The Financial Times. Philip speaks regularly at conferences and other events throughout Europe and North America.

Away from work, Philip's primary leisure activities are canal boats, skiing, playing Bridge (at which he is a Life Master), and dining out.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of Mutable business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Copyright and disclaimer

Bloor Research International Ltd
20–22 Wenlock Road
LONDON N1 7GU
United Kingdom

Tel: **+44 (0)20 7043 9750**
Web: **www.Bloor.eu**
Email: **info@Bloor.eu**