Reference Data Management using Siperian MDM Hub

Technical White Paper
# Table of Contents

**Executive Summary** ........................................................................................................... 2

**A Scaleable Problem** ......................................................................................................... 2

**Reference Data: A Closer Look** .......................................................................................... 5

**Siperian MDM Hub: Manage Your Reference Data More Effectively** ............................... 6

1. Reference Data Creation ..................................................................................................... 6
2. Reference Data Mapping From Various Sources ................................................................. 7
3. Reference Data Workflow and Collaboration .................................................................... 8
4. Hierarchical Reference Data Management ......................................................................... 9
5. Inbound Reference Data Resolution ................................................................................ 10
6. Outbound Reference Data Resolution ............................................................................. 10
7. Reference Data Services .................................................................................................... 10

**Ongoing Monitoring** .......................................................................................................... 11

**Benefits of Siperian MDM Hub for Managing Reference Data** ...................................... 11

**Case Study: Siperian in Action** ....................................................................................... 12

**Glossary of Terms** ........................................................................................................... 14
Executive Summary

Large organizations all over the world recognize that master data management (MDM) has the potential to vastly improve the flow of information across the enterprise, drive more effective system integration efforts and power new business operations. Yet implementing effective MDM solutions often proves more difficult than originally assumed. The challenges posed by integration efforts are diverse, and oftentimes proposed MDM solutions can only address certain aspects of the overall integration initiative. These shortcomings are partially rooted in the approach taken by some MDM solutions providers. More than anything else, however, the success or failure of MDM implementations hinges on technological concerns and one above all others: reference data management.

Reference data management or lookup code management is one of the biggest and most fundamental roadblocks to data integration in general and MDM in particular. Reference data management is a problem that impacts both analytical and operational integration, and it becomes more difficult to solve as the size and scope of an MDM project grows. IT professionals have long tried to address the reference data management problem by manually mapping reference data between structured data sets, or creating new canonical values to match existing fields. These kinds of time-consuming and labor-intensive processes, however, are workable only in small-scale projects integrating two or three systems. For larger companies with dozens or even hundreds of disparate system powering diverse business processes, reference data management challenges all but preclude large-scale data integration projects.

Siperian MDM Hub provides a proven solution for resolving reference data management issues that impact systems integration efforts. An integrated, model-driven and flexible platform, Siperian MDM Hub adapts to business requirements and coexists with all major applications and software packages. These are critical capabilities for large companies supporting diverse, heterogeneous information system environments. Siperian MDM Hub automates critical reference data management processes, enabling you to scale your integration efforts and allow for the seamless addition of any new systems. Its capabilities include reference data creation, reference data mapping from any number of sources, inbound and outbound reference data resolution, hierarchy resolution and related reference data services.

With these powerful features, even the largest global organizations can quickly implement multi-system integration projects and composite applications, and reap the financial and operational advantages of MDM. Moreover, the Hub’s reference data management capabilities simplify complex development tasks, making it possible to initiate valuable new transaction-based business operations. Firms centralizing their efforts to integrate, manage and share reference data also find rapid return on their investment through less manual re-work and, more importantly, reduced regulatory capital reserves for operational risk compliance. In this white paper we will discuss reference data management concepts in more detail, describe the difficulties reference data presents to effective systems integration, and show how Siperian MDM Hub effectively addresses and solves these problems.

A Scaleable Problem

Reference data management in small-scale, point-to-point integration projects is generally not a major concern. Resolving reference data conflicts can be time-consuming and labor-intensive, but is definitely solvable. Where the situation becomes problematic is when organizations need to integrate larger and larger numbers of underlying systems, or incorporate new systems into an existing infrastructure. This is why the reference data problem crops up most often in very large companies that have heterogeneous operations that came about through acquisitions or through rapid organic growth. Within today’s Fortune 500-sized organizations, where it is not at all unusual for more than one hundred customer data sources and systems to be in use, reference data management can be a critical showstopper when it comes to cross-system integration or the creation of new systems relying on divergent data sources. This holds true with both analytical and operational integration.
The real secret sauce of customer data integration is data reliability; you can scrub the data till it shines but it doesn't necessarily mean it's the right data…. companies have been more focused on the business problem than on the underlying quality of the data, specifically the complexities of reconciling hundreds of customer data sources.

Anurag Wadehra, Vice President,

Consider a setting in which a group of marketing analysts from a retail firm want to run a report to discover the overlap of customers across various channels such as in-store sales (POS), web sales and mail order sales. The methodology would require populating a data mart with information from each of these channels, and perhaps augmented with information from the company’s account systems and customer relationship management system. In order for the company's business intelligence applications to slice and dice the data from the various systems, reference data conflicts between all the sources would need to be resolved first. It is impossible to make apples-to-apples comparisons when your business intelligence application is working with fundamentally dissimilar data sets.

The issues surrounding system diversity and proliferation make reference data a concern in operational settings also. Imagine this same retailer plans to open a series of new stores. Before the doors can open the retailer needs to create a whole new set of reference data for the new locations and this information needs to be entered or replicated into all the other systems the company relies on to support critical operations: point of sale, supply chain, ERP and so forth. These are the systems that support just-in-time inventory, real-time transactional processing, and many other processes that drive consumer retail operations today. If the new reference codes aren’t in place when the doors open, the retailer would not be able to process purchases, track inventory, or perform many of the critical IT actions and procedures critical to the business.

A similar scenario prevails in organizations such as banks that must support large customer bases across different lines of business. In most cases the different lines of business operate on different backend systems, i.e. one for checking and savings accounts, another for retirement and yet another for investment banking. When you open an account with a financial institution, your information must be sent to every system that your account data might touch—the account systems, sales force automation systems, call center systems, billing system and so forth. Many large national banks today support diverse legacy systems that built up over time through acquisitions of smaller banks. As such, this process can easily require sharing reference codes across hundreds of systems. Considering the importance of real-time business operations and the growing need for composite systems, it’s clear that an effective reference data management capability is essential for organizations looking to scale integration efforts or quickly add new systems.
Mapping A New Course For Reference Data Management

Commonly referred to as “code tables” or “lookup tables,” reference data provides context for or categorizes data within the database, or even information outside the database. Basic pieces of information such as time zones, geographical information, zip codes and currency are typical reference data, though the category also includes such things as chart of accounts, financial business unit hierarchies, legal entities, accounting master data, financial reference data and class of trade information. To explain the last example in more detail, companies doing business with large numbers of suppliers and vendors often segregate them into separate categories in order to track interactions more closely. The unifying thread with all these classes of information being that reference data creates a detailed framework within which the enterprise can record and understand transactional information as it changes over time. Let’s take a simple example:

<table>
<thead>
<tr>
<th>Source System</th>
<th>Code Value</th>
<th>Code Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRM</td>
<td>US</td>
<td>UNITED STATES OF AMERICA</td>
</tr>
<tr>
<td>ERP</td>
<td>223</td>
<td>USA</td>
</tr>
<tr>
<td>POS</td>
<td>138</td>
<td>UNITED STATES</td>
</tr>
<tr>
<td>CRM</td>
<td>DK</td>
<td>DENMARK, KINGDOM OF</td>
</tr>
<tr>
<td>ERP</td>
<td>335</td>
<td>DENMARK</td>
</tr>
<tr>
<td>POS</td>
<td>7</td>
<td>DANMARK</td>
</tr>
</tbody>
</table>

As we see in Figure 1, while a firm’s CRM system might use one kind of abbreviated country code (US), its ERM system might use a numerical code (223), and its POS system might use an entirely different numerical code (138). Note also that the code description fields use different nomenclature also. While the codes and descriptions all refer to the same basic information, it’s handled differently system to system, and the differing code formats create conflict if and when organizations attempt to integrate the information from their systems. Since MDM depends on the ability to effectively cross reference individual fields in structured data, these codes must either be converted and standardized or otherwise mapped to each other in order to resolve the conflict. The same holds true with other kinds of reference data and related information such as chart of accounts, financial business unit hierarchies, legal entities, accounting master data and foundational financial reference data.
Reference Data: A Closer Look

To gain a deeper understanding of the issues surrounding reference data management, let’s review some basics. The different types of reference data can be segregated into the following categories:

1. Simple Lookups – These are simple codes and associated attributes that can be managed at a single record level. Examples are country codes, prefix codes, region codes, etc.

2. Simple Relationship Lookups – These are simple relationships that define the nature of the connection between two simple or hierarchical lookups, such as person A works for company B or the sales hierarchy for USA Branch.

3. Hierarchical Lookups – This is the set of reference data where the context of a hierarchy brings simple relationships between the instances of entities in the same domain. For examples, products can be organized into a product hierarchy, offices can be organized into a branch hierarchy, people into an organization reporting hierarchy and geographies into a territory hierarchy.

4. Complex Rollups – Complex roll-ups are parent-child relationships between instances of entities in two different domains. For example, people could roll up into branches and branches can then be rolled up into regions and regions then rolled up into operating companies.

Let’s delve a bit deeper into our country code example from the sidebar. As we know, country codes can vary across different systems. But why exactly do these differences cause problems? Here is a list of some of the most common issues that crop up, and scenarios in which the problem must be addressed:

1. The inbound integration point on the hub needs to understand how various values map to each other so that data coming in from one system can easily map to another system. If the code for the geographic entity ‘the United States’ is ‘USA’ in one system and ‘003’ in another, this becomes impossible.

2. The outbound integration points need to understand the reverse logic for the mappings as the data is sent to other participating downstream systems. Therefore, that system interface needs to recognize and understand any data being communicated to downstream systems. Again, a code of ‘USA’ would not be understood by a system that stores the country code information as ‘003’.

3. A firm decides to create a composite application that sits on top of a number of other underlying applications, such as SAP, Siebel, Oracle, mainframe apps etc. This new application needs to be a unifying composite application for the creation or update of critical master data, such as customer, product, etc. The question, then, is which set of reference data from the underlying systems would be the best one to use in the new composite application. Or would it be better to create a set of ‘canonical’ lookup values—a unique new amalgam drawing on the most useful aspects of the reference data from all the underlying systems?

4. You are now moving from integrating two systems to more than a dozen systems. Are there different codes in each of these systems? Where do you maintain and manage the mapping between the different codes across these various systems?

5. Does the addition of every new system require you to break your existing application integration and redo the effort?

6. The country codes were the starting point, but now you also need to manage other enterprise-wide codes for different entities and data. Is the problem spiraling out of control?
Siperian MDM Hub: Manage Your Reference Data More Effectively

As we can see, the problems posed by reference data can cascade into a whole messy pile of IT issues. Siperian's master data management platform provides the features and functionality necessary to resolve even the most imposing reference data management challenges. The Hub's architecture provides unique benefits, including rapid deployment, improved data reliability and trustworthiness, low costs for handling exceptions and ease of data sharing—all leading to faster time-to-value, providing better quality management of data and reporting with superior operational performance. The reality for large IT organizations today is that ripping out legacy systems and replacing them with new integrated systems is simply not a workable or affordable solution. Which is why it is imperative that organizations have the ability to easily integrate installed systems, create composite applications and integrate new systems into current environments—and have a more streamlined process for managing these shared data assets. Siperian MDM Hub gives IT administrators, data stewards and project leaders exactly these capabilities. Specifically, it provides complete control over critical reference data management processes:

1. Reference data creation
2. Reference data mapping from various sources
3. Reference data workflow and collaboration
4. Hierarchical reference data management
5. Inbound reference data resolution
6. Outbound reference data resolution
7. Reference data services

To develop a better understanding of these concepts, let's explore how each comes into play in the context of a master data management implementation. The following sections walk the reader through Siperian MDM Hub's reference data management capabilities, and highlight certain aspects by recounting a real-life implementation at a large manufacturing company. This project established a composite application that serves as the single system of entry for creating and updating all of the firm's global customers. The application was built using Siperian Application Enablement APIs and enables the company to synchronize its single source of truth for customer, supplier and product data to several downstream systems in other lines of business. The business drivers behind the implementation are detailed in the brief case study outline on page 12.

1. Reference Data Creation

Siperian MDM Hub enables both canonical value creation as well as system-specific code creation, using either batch or real-time processing. The administrator first creates a reference data lookup table. This simple table, or base object as it is known within Siperian MDM Hub, contains the unique set of new canonical code values for the firm's new composite application. It also contains columns associated with maintaining the reference data. As this table of canonical values is created in the Siperian user interface (UI), an additional set of underlying tables is also created. These tables store the set of reference values in the underlying systems that map to the canonical values, the history of all changes that are made to the values, and the lineage of how different system values are tied together. The following figure shows the model with the associated cross-reference table for mapping the values across different sources and maintaining the non-unique cross-system values for the lookup codes:
Once canonical values are established in the reference data lookup table, it then becomes possible to configure contributing sources and set cell-level survivorship rules. Put another way, this is the process of determining and setting which aspects of incoming reference data from the other systems will survive as a canonical value. For instance, it might make sense to use country codes from system 1, the time/date codes from system 2 and the account numbering framework from system N to create the canonical, or “best” values. With Siperian MDM Hub, it is possible to easily identify and manage what sources will contribute to the creation of the new values.

2. Reference Data Mapping From Various Sources

The next step in managing reference data for composite applications is to resolve different reference data values to a set of canonical best values by defining mapping relationships between the different systems. This can be done for each system using the Mapping functionality in Siperian. In the case of our manufacturing company, the administrator created and configured the mappings for each source. Once the Base Objects were defined, it then became possible to start mapping the input from various sources. This process also defined the batch and the real-time interfaces for loading and updating the reference data into Siperian.
3. Reference Data Workflow and Collaboration

Creating a “single source of truth” necessarily involves stakeholders from across the enterprise, and requires input from and agreement among many line of business owners, system administrators and data stewards. To facilitate this process—and the actual loading of data into the Hub—Siperian enables users to provide input on what the canonical values should be, and also identify how the different source system values map to these canonical values. The language frames used by the various source systems can and should be reflected in the composite application. In this way, line of business owners and the administrators of the various systems can ensure that the canonical values reflect their business priorities and the nomenclature used by system end users. This information can easily be captured and shared using spreadsheets or by creating collaboration processes using a BPM tool. This process also facilitates the initial load of the reference data into the Hub. Here is a sample of how this information can be captured:

<table>
<thead>
<tr>
<th>Code Type</th>
<th>Source System</th>
<th>Code Value from Source System</th>
<th>Code Description from Source System</th>
<th>Canonical Code Value in Siperian</th>
<th>Canonical Code Description in Siperian</th>
</tr>
</thead>
<tbody>
<tr>
<td>State Code</td>
<td>CRM</td>
<td>“Calif.”</td>
<td>“California”</td>
<td>CA</td>
<td>CALIFORNIA</td>
</tr>
<tr>
<td>State Code</td>
<td>CRM</td>
<td>“Colo.”</td>
<td>“Colorado”</td>
<td>CO</td>
<td>COLORADO</td>
</tr>
<tr>
<td>State Code</td>
<td>CRM</td>
<td>“Conn.”</td>
<td>“Connecticut”</td>
<td>CT</td>
<td>CONNETICUT</td>
</tr>
<tr>
<td>State Code</td>
<td>CRM</td>
<td>“Del.”</td>
<td>“Delaware”</td>
<td>DE</td>
<td>DELAWARE</td>
</tr>
<tr>
<td>State Code</td>
<td>CRM</td>
<td>“D.C.”</td>
<td>“Washington D.C.”</td>
<td>DC</td>
<td>DISTRICT OF COLUMBIA</td>
</tr>
</tbody>
</table>

Figure 4 - Mapping Source System values to the Canonical Siperian values
Once this data has been captured with the help of business users, it can then be loaded into Siperian using the mappings for the reference data elements from each source. When this process is complete, the data will not only represent the best or canonical values for data across the organization, but it will also represent the various source system mappings or cross-reference records.

At this point, the clean reference data can now be connected to the master data as it is loaded into the Hub (batch or real-time). The connection between the master data and reference data can be configured for any master data entity inside Siperian. At runtime this metadata is used to resolve all the lookup values to the correct canonical values for storage in Siperian. In the example in Figure 4 above, the state-level source codes from the manufacturer’s CRM system—which were originally stored using antiquated postal codes—is mapped to new canonical values. Siperian not only manages this translation but also keeps track of the correct lineage at a master data cross-reference level for each resolved and original source key value. The Siperian system can also be used to create and maintain hierarchical reference data or hierarchical lookups, for example the Sales Territory, Chart of Accounts (COA) or Region hierarchies. Hierarchy Manager not only allows you to model these but also allows you to create and maintain this Hierarchical Lookup data through the UI.

4. Hierarchical Reference Data Management

Siperian’s powerful mapping and workflow capabilities provide fine-grained control over reference data such as territories, organizational structures, etc. This can be done using Siperian’s Hierarchy Management functionality—which enables the user to identify relationships between data objects such as spouse, parent and child—and related batch and real-time capabilities. This functionality enables end users to navigate from one hierarchy to another, pivoting between them via entities common to both. In this way, a user can view all the relationships a given entity has, across all the available source systems, in a single view.

Assume our manufacturing company tracks customers by multiple designations across its various systems. For instance, its product system groups them by sales level or purchasing amount, while the CRM system might rely on a hierarchy of geographic locations, such as domestic or export and then sub-designations beneath that. A financial analyst wishing to understand the company’s export business in the context of sales performance by customer could run a simple search of the
Siperian database of relationships to immediately generate a single view, similar to what is shown in Figure 5. The relationships illustrated have been gathered from all the various underlying systems, each of which contributes to the organization’s total information about its export business. In the context of the composite application used by the manufacturer, this ability to immediately see the full scope of its customer relationships is extremely useful in a) eliminating confusion over which facility can best meet customer needs and b) enabling the company to optimize its sales and marketing efforts.

5. Inbound Reference Data Resolution
Once reference data has been mapped, hierarchical relationships have been established and lookup rules set, Siperian MDM Hub can automatically resolve virtually any number of data feeds to the newly established canonical values. This is the critical capability that makes it possible to create composite applications for any number of data sources across the enterprise. Moreover, the Siperian system can accommodate integration processes not only for lookup data but customer or product data as well. As that data changes on the source systems or as new data is created, that data flows into Siperian. At that point, Siperian automatically maps the new data to the reference data, thereby creating and maintaining a fully accurate, single view—of customer data, vendor data, product data, etc.

6. Outbound Reference Data Resolution
In the same way the Hub supports automatic inbound data resolution, it also supports outbound data resolution. In reference data mapping (or reverse mapping) for outbound data, the Hub automatically captures all data that supports reverse lookups. So, once the data is reconciled in Siperian, if any new customer data is created in, say for instance the manufacturer’s sales tracking system, Siperian sends out messages to the various other systems within the company—ERP, CRM, content management, etc.—alerting them to the new data, and automatically providing these systems with the correct lookup codes by utilizing the lookup mappings parameters stored in Siperian. In this way, all systems across the company can be kept in synch with each other.

7. Reference Data Services
Siperian MDM Hub provides a range of tools, features and services that can be used for the orchestration of data to various downstream systems and for ongoing reference data management. These services can be used for providing reference data for building composite applications, especially the areas such as drop-down lists, which can present a major obstacle when building composite apps. In addition, the Hub can be used to expose the underlying reference data for use in integration through various interfaces. These are the APIs provided by Siperian or services created using the APIs, and they can be divided into two categories:

- **Lookup Services for Reference Data**
  The user can define a set of lookup service interfaces to expose the underlying data captured in the Hub, and manage it according to whatever type of integration is being performed. These are generic interfaces that can be used not only in data synchronization scenarios but also in composite application scenarios or data federation scenarios. They also enable the quick resolution of any conflicts within reference data by relying on the reference data information and the corresponding mappings for each source as provided by Siperian.

- **Authoring Services for Reference Data**
  Authoring services are based on the Siperian APIs for inserting or updating data for the specific reference data objects. These APIs are both generic and schema-specific (strongly typed and based on the reference data object definition). Schema-specific APIs for the reference data objects can be generated once the reference data model is defined using the Service Integration Framework (SIF).
Ongoing Monitoring

Monitoring and sharing changes to reference data are critical capabilities that make any reference data implementation successful. Siperian enables users to monitor Hub activities for any critical changes to reference data, and then publish these changes to a downstream system or interfaces for data propagation or notification. It is easy to configure the Event Management rules in order to watch for any critical changes that take place around the reference data either inside the Hub or in any contributing source system. Siperian can also be used to enable collaborative processes for maintaining reference data on an ongoing basis. You can do so using State Management, Event Management and Workload Management, all of which are provided out of the box. Moreover, collaborative processes for managing reference data can be enabled in a variety of ways.

Option 1 – Inside Siperian
This option allows the users to leverage the basic workload and state management capabilities in Siperian to manage the creation and management of reference data. This can be done using either the Data Manager, Merge Manager and Hierarchy Manager GUIs provided as a part of the Siperian toolkit, or by leveraging the APIs to expose the same functionality in a custom UI.

Option 2 – Using a BPM tool
In order to enable collaborative processes, the second option is to use a BPM tool in conjunction with Siperian to define robust and cross-organization Reference Data Governance processes. This can, once again, be done by leveraging the Siperian APIs and capabilities for the following areas:

- Create and Update Management – APIs that allow users/business processes to create or update reference data.
- Event Management – Define rules in Siperian to monitor the critical reference data changes and the subsequent routing of these events to downstream systems.
- State Management – Allows users to capture in-process data from a workflow and manage it as a part of the reference data.
- Workload Management – Provides the ability to manage the distribution of work across the various users responsible for the task of managing reference data.

Benefits of Siperian MDM Hub for Managing Reference Data

Only Siperian MDM Hub has a completely integrated, model-driven and flexible architecture that allows your organization to resolve any of the most pressing reference data management challenges facing enterprises today. Siperian MDM Hub is faster to deploy, easier to manage, and simpler to migrate over time for all reference data management and master data management projects. Siperian’s integrated architecture and robust feature set enables easy reference data creation, mapping and hierarchy resolution. With these powerful features, even the largest global organizations can quickly implement multi-system integration projects and composite applications.

Faster and More Cost Effective
In fact, because Siperian MDM Hub’s reference data management capabilities simplify so many complex development tasks, incorporating new systems and initiating valuable new transaction-based business operations can be done faster and at lower cost than ever before. And, once reference data is normalized and available in Siperian MDM Hub, it can be leveraged for any number or type of integration or reporting scenarios. Some of the scenarios that can make use of this data are:

- Reporting and Analytics
  Leverage the reference data in your reporting environment to gain more comprehensive and more accurate visibility into critical business processes. It is easy to do by extracting data from the Hub, which provides access to the underlying data model and the associated reference data.

- Composite Applications
  Users can combine different applications or mash-ups to create valuable new operations or services. Because Siperian MDM Hub allows for the mapping of reference data values—not only
for the incoming data but also for the data that is changed or created through the composite application—deploying new offering is quick and easy.

**Data Synchronization**

Many data synchronization initiatives based on replication of master data to multiple systems run aground due to the differences in underlying reference data. Siperian MDM Hub enables you to easily overcome these issues. The Hub’s scalable integration architectures can be defined by creating reference data services that can provide “to” and “from” mapping, thereby resolving integration problems before they can arise.

**Case Study: Siperian in Action**

Many Siperian customers initiating an integration effort start the process by resolving reference data issues in order to quickly deliver value to their users within a short period of time. To illustrate how this is done, let’s consider a case involving a large manufacturing company using Siperian MDM Hub to implement a master data management project. This provider of paper and paper products to global industries and printing services providers had grown steadily over the years, acquiring and incorporating many smaller operations along the way. This growth led to challenges in many areas, including its vendor-side operations, customer-side operations and technical/IT operations.

**Business Challenges**

On the vendor side, the company was hobbled by poor contract management and pricing controls. It offered a bewildering array of discounts based on often confusing or even arbitrary rules, and had difficulty applying these vendor discounts at the enterprise level. It also lacked insight into vendor and supplier data across its manufacturing facilities. The end result was that the company often overpaid on contracts, suffered from poor inventory management and failed to take full advantage of available pricing discounts.

On the customer-side, the fundamental lack of visibility into its customer base sometime resulted in sister facilities within the company competing inadvertently for business from the same entity. These problems extended across the enterprise, undermining overall sales and marketing efforts, manifesting in poor rebate management, and difficulty in enforcing uniform pricing policies. Adding to these problems was the fact that the manufacturer had an extremely diverse IT infrastructure that had built up over years of inorganic growth. Integrating new systems into the enterprise within a short period of time was extremely difficult, and often not possible.

**Technical Challenges**

To address these business problems, company managers resolved to create three new composite applications to better manage its vendor-side operations, customer-side operations and product management operations. But first, the manufacturer had to solve its reference data management challenges. The primary reference data management challenge was to normalize the reference data across the three core transactional systems (1, 2 and N) so that the new composite applications would be able to integrate individual customer account information, vendor/supply data and product data as needed. The question was, what would be the optimal method for reconciling divergent reference data? Adopt the codes and hierarchies from one of the systems and map them to the codes in the other two? In this case, it was decided that the best solution would be to create a set of ‘canonical’ lookup values that the composite application could use to map to and understand reference data as it flowed from the three underlying systems. Accordingly, the reference data management phase of the manufacturing firm’s master data management implementation mapped closely to the seven-point process outlined previously in this paper.

**Solution**

Using Siperian MDM Hub, the company was able to create composite applications to provide
single views of both its vendor-side operations and its customer operations, and create a new product data management system. Dubbed Global Customer On-boarding, the firm’s new customer-facing system uses optimized data governance policies and procedures to centralize its customer management, sales and marketing operations. The company used Siperian Application Enablement APIs to quickly integrate its many customer-facing applications, and it now uses this system for creating and updating all global customer data. With this new single source of truth residing in the Hub, the manufacturer is able to synchronize clean, correct customer information with downstream systems in other lines of business.

The company went through the same process with its vendor-side systems, creating a Vendor On-boarding system that has eliminated conflicts in its vendor-side data; and also used the Hub to create a new product data management system as well. This -application is the new system of record for product data, and ongoing creation of new product data and changes to existing product data is done using an application built on top of Siperian. All updates to the product data will be sent to downstream systems, as is done with customer and vendor data.

The diagram below illustrates in more detail how Siperian MDM Hub enabled the manufacturer to resolve its reference data management challenges. In the project, data from multiple existing systems (Systems 1, 2 and N, or CRM, ERP and POS, respectively) is integrated through the Siberian Hub to populate a composite application. The diagram provides a rough representation of the systems involved and the flow of information involved.

**Benefits**

By using Siperian MDM Hub to address its problems with conflicted customer, supply and product data, the manufacturer was able to solve many of the issues facing its business. With a single view of vendor data and new insights into purchasing trends, the company was able to improve its on-contract purchasing and take advantage of a far higher percentage of available pricing discounts. As such, its supply-side costs were reduced significantly. Similarly on the customer side, company-wide visibility meant that sales and marketing managers were able to eliminate duplicative outreach efforts and redundant spending. With a more intelligent and targeted approach, the company was able to improve the performance of its customer-side operations even while it reduced spending. The improvements extended to its product-focused operations as well, enabling the company to retire some of its older systems, saving on IT costs. Overall, the project was such a success that the company plans to extend its Siperian MDM Hub implementation for employee management operations also.
Glossary of Terms

Master Data Management (MDM)
Master Data Management is the controlled process by which the master data is created and maintained as the system of record for the enterprise. MDM is implemented in order to ensure that the master data is validated as correct, consistent and complete and optionally circulated in context for consumption by internal or external business processes, applications or users. Ultimately MDM is deployed as part of the broader Data Governance program that involves a combination of technology, people, policy and process.

Master Data
Master Data is a collection of common, core entities, their attributes and their values that are considered critical to a company's business and are required for use in two or more systems or business processes. Examples of master data include customer, product, employee, supplier, location, etc. Complexity arises from the fact that master data is often strewn across many channels and applications within an organization and invariably contain duplicate and conflicting data.

Reference Data
Reference data is any kind of data that is used solely to categorize other data found in a database, or solely for relating data in a database to information beyond the boundaries of the enterprise. As a category, reference data is similar to, though distinct from, master data, and would include pieces of information such as time zones and geographical information, chart of accounts, financial business unit hierarchies, legal entities, accounting master data, financial reference data and class of trade information. (Reference data is also referred to as look-up data, or look-up codes.)

Canonical Value
In reference data management, a standard value assigned to a data object that can be mapped to identical or equivalent data objects in other systems, or recognized by other systems as the same.

Survivorship Rule
A guideline used in the process of determining and setting which aspects of incoming reference data from other systems will survive as a canonical value. Particular aspects of reference data from a variety of systems can be used to create the canonical or “best” values. Survivorship rules govern which data elements are retained and which are eliminated from the final “best” value.

Code Table
A table within a database containing reference data or look-up codes, along with the metadata having to do with the ongoing management of the reference data.

Composite Application
An application that draws on or integrates a variety of data sources within the enterprise to provide visibility into or capabilities across, various lines of business, disciplines or functions.