

# The information perspective of SOA design, Part 2: The value of applying the business glossary pattern in SOA

Skill Level: Intermediate

[Brian Byrne \(byrneb@us.ibm.com\)](mailto:byrneb@us.ibm.com)  
Industry Models and Integration Architect  
IBM

[John Kling \(jkling@us.ibm.com\)](mailto:jkling@us.ibm.com)  
Consulting and Services Architect  
IBM

[Dr. Guenter Sauter \(gsauter@us.ibm.com\)](mailto:gsauter@us.ibm.com)  
Senior IT Architect and Manager  
IBM

[Peter Worcester \(pworcest@us.ibm.com\)](mailto:pworcest@us.ibm.com)  
Services Solution Marketing Manager  
IBM

14 Feb 2008

Do you find it challenging when key business terms cause confusion, back and forth debates over what they (should) mean, delays, late changes, or even complete failure in your SOA or data integration projects? This second article in the series "The information perspective of SOA design" helps you eliminate these misunderstandings by introducing the concept of a business glossary. Discover the value of a business glossary in SOA and learn how to define and use it to communicate more clearly with your colleagues.

## Introduction

[Read all the articles in this series](#)

1. [Introduction to the information perspective of an SOA](#)
2. [The value of applying the business glossary pattern in SOA](#)
3. [Use the IBM WebSphere business glossary in SOA design](#)
4. [The value of applying the canonical modeling pattern in SOA](#)
5. [Use of Rational Data Architect in SOA](#)
6. [The value of applying the data quality analysis pattern in SOA](#)
7. [The execution approach for the data quality analysis pattern in SOA](#)
8. [Use of IBM WebSphere Information Analyzer in SOA design](#)

From the creation of a Service Oriented Architecture (SOA) or data integration engagement and throughout its duration, there is often no common business glossary that defines the terms related to processes, services, and data. A common definition of business terms is essential to eliminate the ambiguity that complicates many data integration initiatives. Without agreement of what we mean by a "customer", "member", and so on, you cannot correctly implement services related to those concepts, or even ensure that all stakeholders are in agreement about the data that makes up these concepts. It is critical to have a common understanding between business analysts and the technical community on the terminology used across processes, services and data, including the semantics, structure and format of data structures.

This article addresses the need for a business glossary within the context of an SOA. It shows you how to define and use this helpful glossary.

## Motivation and problem statement

Unfortunately, it is all too common that the same term is interpreted in different ways between business and IT experts and other various lines of business. Many times, businesses have redundant copies of seemingly similar or identical data elements, whose context is only valid within the data store in which it resides. Often when this element is exposed to outside consumers, the information loses its context, or, worse, it becomes misleading or invalid. Therein lies the fundamental value of a business glossary. It is the contract that aligns the definition of data elements so that its context is meaningful for all consumers of this data.

A business glossary should contain not only the agreed-upon definition for a data element, but any variations or dependencies associated with that element. This eventually helps drive the definition of conceptual and logical data models. Ultimately this enables the components and participants in an SOA solution to arrive at an aligned and consistent definition of business information in context.

In some scenarios where a service requires information from several sources, there may be redundant data elements with the same name or an implied definition. The definitions or contexts of these data elements may be quite different and have meaning only within their particular data store, possibly only when retrieved by their associated program. For these cases, the business glossary helps identify and

resolve such conflicts by providing one consistent and common business definition and dependencies for each entity across the enterprise. When users from different segments talk about "customer" or "revenue" everyone will know exactly what they are referring to.

The possible impacts of not specifying a business glossary are:

- Risk of missing fundamental information requirements of the business
- Increased cost in moving forward with projects where the information requirements are not fully understood
- Increased time spent reworking unclear or misunderstood requirements

In the context of SOA, surfacing these common business definitions is the key to facilitating discussions, not just about the semantics of data elements, but also about the grouping of data that is most reusable across multiple business domains. For example, it is relatively easy to get multiple lines of business to agree that they need a service to access customer details. What is considerably more difficult is getting agreement across these lines of business about which data elements are relevant for the business concept of a customer. Having achieved this, it is still more of a challenge to identify which of these data elements make up a reusable customer data aggregate, and which are ancillary elements that are retrieved through subsequent service invocations. Without a business glossary, these discussions usually become embroiled in debate over naming and business terminology. Worse still, these discussions may never arise in the first place because a lack of precision in business terms hides the fact that there is disagreement in the first place.

In the case where the scope of the engagement is an enterprise master data management (MDM), a similar problem exists. Multiple source systems feeding the MDM repository play the role of the multiple lines of business above -- each with a distinct interpretation of the semantics, structure and aggregation of master data. Without a clear understanding of the variances in the semantics of these data sets, it is difficult to map them into a single master data repository. Again, a business glossary can reduce the ambiguity in this context, thus enabling a single set of business definitions across multiple source systems.

Similar patterns exist within data federation and data consolidation. In these cases, the requirement is much the same: to understand the meaning of data in business terms, thus ensuring that resulting solutions meet the original business objectives for the project.

## A case study

An IBM client in the health care industry struggles with a problem that is common in many engagements: every meeting opens with a dispute over whose numbers or

reports are correct. The problem is not that this company needs a data warehouse. In fact, they have several data warehouses. But all of that information creates more confusion in reporting. For example, the executive management can never get the functional areas of the business such as the Medical Management department and the Sales department to agree on the number of members serviced by the health maintenance organization (HMO).

This company contacts IBM to help them understand the root cause of the problem and to suggest remedies. While data warehouses are used to centralize data from operational systems, there is no agreed-upon business definition as to the meaning of "member". To the Medical Management department, a member is a potential patient, the subscriber and all dependents eligible for care. To the Sales department, a member is a subscriber and all dependents if that subscriber is eligible for renewal. Dependents of a deceased subscriber are excluded from their counts. This means that the member counts for Medical Management and Sales are the same for one plan (such as, a local HMO) but are different for another. The differences vary by year and are never be the same. For executives, all of this is maddening because they feel they can't believe their own reporting.

A common business glossary effectively addresses this problem because it establishes terms that have a single and commonly agreed-upon business meaning. All reports can then use the agreed-upon terms consistently across reporting contexts.

## The concept of a business glossary

A business glossary, sometimes referred to as a data dictionary, is the artifact that defines the terms and data associated with an initiative. Depending on the extent and type of the engagement, the scope of a business glossary can vary, defining terms within the context of a (product or line of business) silo, an information domain, or the enterprise. The preferable scenario is where business terms are defined across the context of the entire enterprise, driving consistency of business terms across all projects.

It is, however, common for different departments or lines of business to have different semantic contexts for what would seem to be the same term. For example, consider what the meaning of the element "address" is to a distribution department. This is likely to be the "ship to" address. However, to the accounting department, its meaning is most likely to be a "bill to" address. To sales and marketing, its meaning will likely be a "call on" or "contact" address. This is a very simplified example and is easily dealt with using a name prefix or having three different address fields. Nevertheless, there needs to be a way to document and identify which type of address we are dealing with, and what each one means.

The business glossary defines the language of the business and, by extension, the

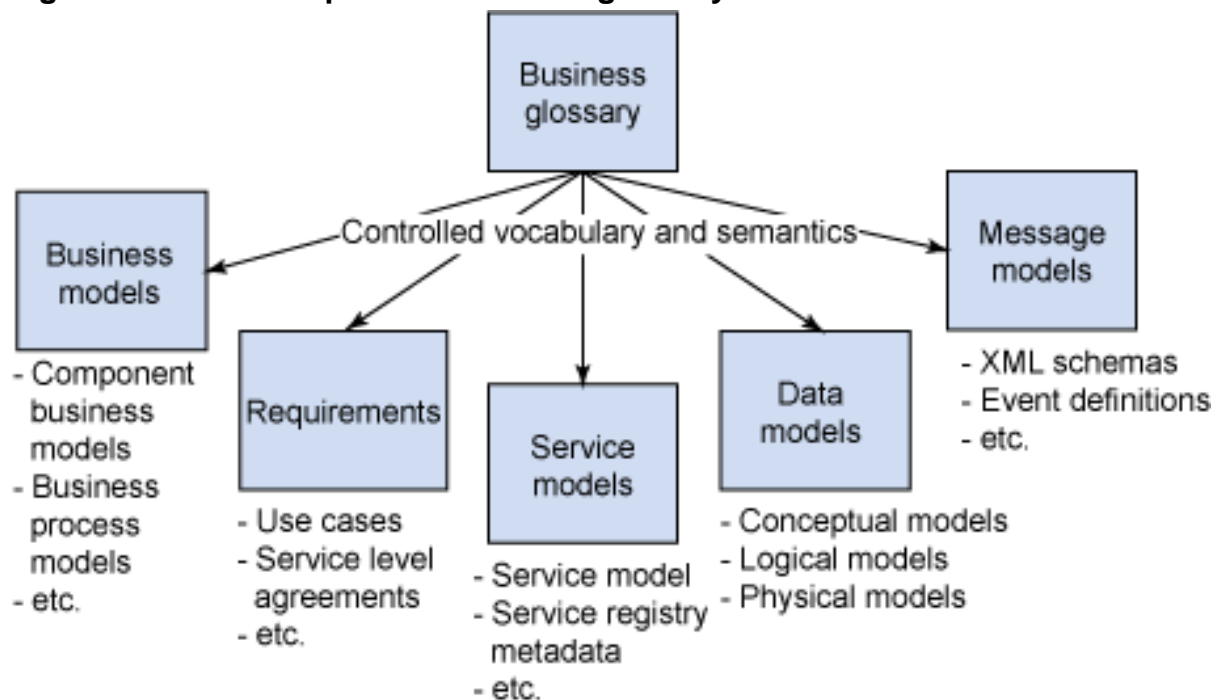
language of the project. Therefore, care needs to be exercised that the terms defined in the business glossary are fully qualified and that specific descriptive definitions are provided. To the extent possible, a definition that applies enterprise-wide should be crafted. Where departments use a term differently, those definitions should be captured and associated with their appropriate contexts (department).

When an organization builds an enterprise-wide business glossary, it may include both semantic and representational definitions for terms. The *semantic definitions* focus on creating a precise meaning for each term. *Representational definitions* focus on how each term is represented in an IT system such as an integer, string or date format (see data type). Business glossaries are one step along a pathway of creating precise semantic and representational definitions for an organization.

In any SOA or data integration initiative, the business glossary captures terms that surface during any of the discovery activities, such as process decomposition, reuse analysis or analysis of existing assets. Terms can be related to process activities, business goals or can consist only of the definitions of the identified individual sources.

This results in a glossary model that maps to the artifacts that emerge during the various forms of structured analysis - business models, data models, requirements models, and so on. Figure 1 shows the relationship of the business glossary to the other artifacts:

**Figure 1. Relationship of the business glossary to other artifacts**



Assuming a glossary does not exist with an organization, the question arises of how to go about building one. In fact, the pattern is very similar whether no glossaries exist, or whether multiple fragmented glossaries exist across the enterprise.

### **Who creates a business glossary?**

This leads to the discussion of the roles necessary to create a proper business glossary. In some organizations there are existing business analysts who understand the business definitions of the data in question. In other organizations, there are informal experts who are the historical, informal stewards of data. In many cases, there is lack of formal definitions and dependencies associated with most of a company's data. More and more frequently, there is an emerging role in organizations called the data steward. This is typically the role that most often manages the creation and maintenance of the business glossary.

The owner(s) of the business glossary varies from organization to organization and even within silos in the same organization. Ideally, information domains and their ownership should be defined within an enterprise data/IT governance structure, and these same information domains and ownership hierarchies can be applied to the business glossary. If such governance structures are not already in place then it is likely that the data architect will play this stewardship role for the duration of the SOA project with a view to identifying the long term ownership strategy by the project's end. It is highly recommended that any project contains or utilizes a governance process. If such a process does not already exist within a company, then the project should include implementing such a governance process.

Typically, there is a business analyst or data steward identified for each information domain or perhaps, at an even more granular level, for each operational data source, domain, or entity involved in the solution. In some instances that could be the same person, but in other cases, it may be individuals from the LOB or segment that owns the data involved. It is even possible that any one source may have multiple data stewards for the data, each with expertise in particular subject matter. It is possible that there will be more than one steward for any one term. Take, for example, the term "customer type." There may be customers from marketing or finance, and set of customer-related data may have a data steward from that department or functional area. In the example above, "Address" may require adding a qualifier or extending the data structure to identify the category of "address". The data architect will be helpful in suggesting ways to logically associate and identify the type of "address" being asked for. This is just one more reason to have a mixture of skill sets helping to drive these definitions. If left solely to a business analyst, then the resulting view may not meet the needs of all the dependent downstream activities in the subsequent methodology phases.

An appropriate subject matter expert (SME) must be identified and assigned the role of data steward for any particular source or portion of a source. This should be someone who understands the business use of all the candidate business terms

identified for potential inclusion in the project. This may or may not be a single individual depending on the amount of identified source data and the individual's knowledge of that data. They should also know, or be able to learn, the dependencies and relationships between all these entities.

It is also often very helpful to have a data architect available and involved in this process as they typically understand the physical constraints and structural aspects of the data sources. They are also helpful in determining the relationships and dependencies among the data.

The notion of the role of SME, business analyst and data steward are all valid, however it means that these individuals have the arduous task of seeking out the business experts who have the actual business definitions of the terms and getting agreement from everyone involved on the final definition of that term. This is not a task to be taken lightly and often will take a tremendous amount of time and effort to achieve. This project is not just taking on the definition of its core informational requirements, but is seeking to define these terms across the context of the whole enterprise. As mentioned earlier, this is the contractual agreement across the enterprise of the real definition of each element or term. This establishes the common language upon which every consumer of data relies. The mechanics of achieving this vary from company to company and project to project.

### **When is a business glossary created?**

You have to remember that a business glossary cannot be started early enough, and it is not just an exercise initiated in the early discovery phases. A business glossary can and should be considered as early as possible -- even during the requirements-gathering and project-definition phases. A business glossary is not limited to existing data stores or databases; it also contains the definitions for all business terms used to describe business processes and services in an SOA. The earlier the glossary is being developed, the sooner a foundation for consistency of terminology throughout the project and the enterprise is accomplished.

The business glossary should be established and developed during any initial discovery phase, regardless of the specific methodology being applied. As the project progresses through its various phases, the business glossary is continually updated and refined.

### **How is a business glossary created?**

In constructing a business glossary, the following steps should be considered.

1. **Gathering information sources**  
Business terms for inclusion in a glossary may come from a range of sources, for example industry standards, or IBM's Industry Models. Other common sources include requirements documents, existing data

dictionaries, and even legacy solutions.

There may be an existing business glossary for one or more existing systems or one may be defined by an industry standard. If such a glossary exists, then it can be reviewed and included for integration into the common business glossary.

## 2. **Extracting business terms**

The most valuable aspect to the business glossary is the hardest to obtain -- that is an agreed-upon, universal definition of a business concept. This can be acquired by interviewing subject matter experts, through facilitated sessions or questionnaires. The breadth of scope across which a term is used is usually proportional to the difficulty in achieving agreement on its meaning. Where a term is used only within a narrow scope of the business, then the SMEs or business analysts may already have the agreed-upon definition. In the case where a term is going to be used by the entire enterprise, then reaching an agreement on the definition can become measurably harder and might require first gathering the individual definitions and then having the individual SME meet and try to come to a common understanding and definition. Often this results in what was initially a single term being fragmented into multiple terms to satisfy all contexts. This does not necessarily represent duplication within the glossary. Often the same term is adopted by multiple stakeholders to mean very different things -- a situation which persists because of the ambiguity that occurs where there is no clear business definition. Under investigation, the fact that there are multiple different requirements being described using the same words becomes clear, and providing detailed business definitions in the form of a glossary acts as a catalyst to differentiate these requirements.

One simple aspect of the business glossary that is often overlooked and which is the easiest information to obtain is the physical aspects of terms (that is, data-type definitions, constraints, and so on). It is worth noting that, depending on the term, the physical structure may not be a mandatory aspect to collect and may be defined later or in the specification phase as data models are developed. However, it is worth maintaining that information in the business glossary even if it is added at a later date. This can be done manually or using some of the automated information analysis tools available on the market. These will help discover the physical characteristics of terms and can be very helpful in assessing the existing quality of the information.

## 3. **Building a glossary**

It is important to maintain integrity and normalization within the glossary.

Allowing a glossary to grow organically can result in significant overlap of business terms, which further adds to the confusion in terminology that the glossary is seeking to eliminate. When adding terms to a business glossary, the terms should be reviewed in order to determine their overlap with existing terms, and adjusted accordingly to avoid redundancy and other conflicts within the glossary. However, these conflicts may be retained as aliases for the agreed upon term. When this is done, the context of the alias should be noted which allows local or parochial usages to be tracked. This avoids confusion when dealing with groups who insist on using their local usages.

#### 4. **Validation**

There is a number of criteria that indicate a business glossary is nearing completion. For example, terminology can be seen to be convergent -- fewer and fewer new business definitions are being identified across additional domains. A second indicator is that all key client personnel have validated the completeness of the sources used to build business language definitions, and all their terms have been successfully classified by business concepts. Quality of business definitions is also a critical checkpoint. A business stakeholder should be able to validate that the terms are meaningful to the business, are properly classified, and have the proper context.

A well-formed business glossary can form a valuable input into canonical data modeling activities, providing the core definitions required to establish entities and relationships within these models. Although the business glossary will be an input into the canonical data modeling activities, it is important to realize that a business glossary does not replace a canonical data model. There is a significant difference in the degree of detail, formalization, and, of course, structure. The role of a business glossary is to provide clear business terminology. A logical data model takes this a step farther, analyzing the detailed structure of the data, including relationships, sub-typing, attribution and containment.

#### **Updating the business glossary**

The business glossary is not a static document that is created one time and then used as reference for the initiative. The business glossary is a living, iterative artifact. Whether it is a document or a glossary maintained in a tool such as WebSphere Business Glossary, it is intended to be a constantly revised and maturing artifact.

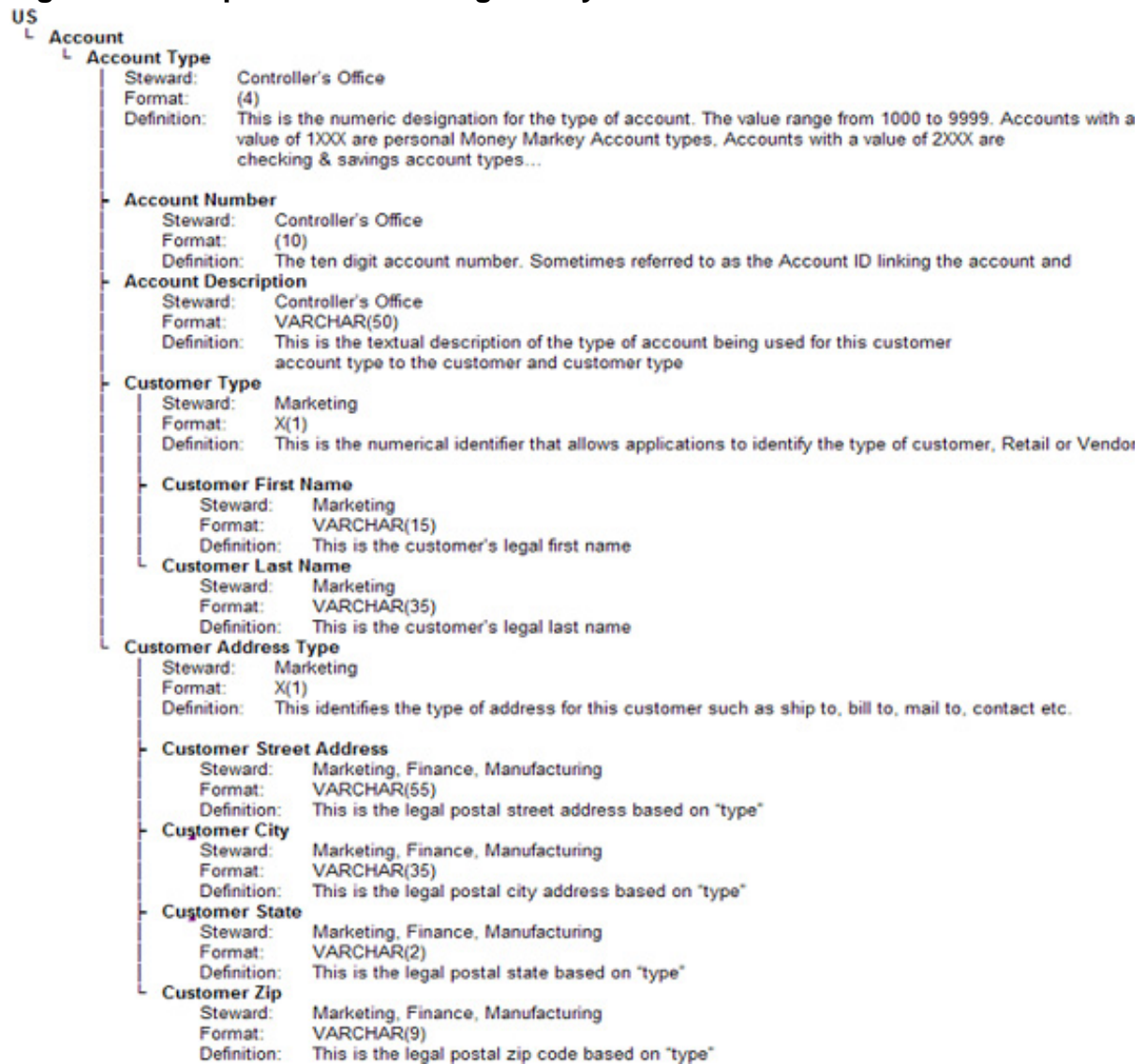
As the initiative progresses from the initial discovery phase to later phases the document becomes more mature, increasingly becoming the accepted point of reference for business terms. The basic physical structure of the artifact does not

change, whether or not the artifact is in document form or captured in a tool.

## Example

Below is an example from a well-defined business glossary centered around an account-opening procedure. Depending on the maturity of the existing definitions, (for example, early in the identification phase) there may or may not be full definitions and values for all terms. The glossary matures as the project matures through the various methodology phases. As more is learned about the business terms, more information should be updated in the business glossary.

**Figure 2. Example of a business glossary**



## Conclusion

In conclusion, the business glossary is the definitive artifact that controls and defines the common vocabulary and, therefore, the semantics of terms and related taxonomy. It is an important starting point in data integration as well as in SOA initiatives to ensure that various roles in business and IT across the organization have the same understanding, not just of which terms are which, but what terms come together in an SOA context to form reusable information structures. No initiative is likely to deliver trusted information if it is unclear what information to deliver, or what the meaning of that information is.



# Resources

## Learn

- Check out the rest of this [series](#) to learn more about topics that were introduced in this article.

## Get products and technologies

- Create, manage & share an enterprise vocabulary and classification system with [IBM Information Server](#) and in particular [WebSphere Business Glossary](#).
- Accelerate projects, reduce risk, and leverage extensive industry-specific glossaries with [IBM Industry Models](#).

## Discuss

- Participate in [developerWorks blogs](#) and get involved in the developerWorks community.

## About the authors

### Brian Byrne

Brian Byrne has over 10 years experience in the design and development of distributed systems, spending 7 years driving the architecture of Industry Models across a range of industries. Brian is currently an architect within IBM's Information Management organization.

---

### John Kling

John Kling is an architect in the Information Services Practice within IBM's Global Business Services. He is responsible for leading large client engagements that focus on data quality, data integration and master data management. He is currently the data team lead for the SAP implementation of a Fortune 500 industrial company.

---

### Dr. Guenter Sauter

Guenter Sauter is an architect in the Information Platform & Solutions segment within IBM's software group. He is driving architectural patterns and usage scenarios across IBM's master data management and information platform technologies. Until recently,

he was the head of an architect team developing the architecture approach, patterns and best practices for Information as a Service. He is the technical co-lead for IBM's SOA Scenario on Information as a Service.

---

#### Peter Worcester

Peter joined IBM three years ago after almost 25 years at institutions like the US Dept. of Defense, GE Corporate and Morgan Stanley where he held technical leadership positions and gained valuable experience in Enterprise Architecture and Enterprise Data Integration. He initially joined IBM as a Sr. IT Architect as part of the architect team for Information as a Service. Currently he is a Solutions Marketing Manager for the IPS Global Services organization, specializing in MDM solutions.