As the amount of data an enterprise manages continues to grow, an organization must contend with two unique challenges: the sheer volume of that data and the multiple methods used to standardize this information across systems. These challenges often lead to a disparate, disconnected information technology (IT) environment, where the information in any system can operate in an isolated silo, creating multiple “versions of the truth.” In this article, the authors examine how one organization met its data management challenges by analyzing the data it receives, creating repeatable routines to improve the quality of that data, and implementing controls to maintain high levels of data integrity. Through this combination of process improvement and technology, organizations can establish a more team-based view of data that spans the IT and line-of-business environments—and create a quality-focused data management culture.

Key words
best practices, case study, data governance, data management, data profiling, data quality

INTRODUCTION

Many companies have come to realize the substantial impact underlying data can have on every aspect of an organization. For the past 20 or 30 years, companies have purchased billions of dollars worth of applications designed to maximize the data they have in hopes of providing a data management structure for that information on an ongoing basis. These business applications are often designed to augment or administrate a business process or processes (for example, a customer relationship management (CRM) system manages sales and marketing activities). From that standpoint, much of the work on these applications has focused on just that—the processes that feed into a “data supply chain” (Dyché and Levy 2006, 12).

However, an organization’s ability to make effective decisions is based largely on the data in front of decision makers. From executive-level decisions about mergers and acquisition (M&A) activity to a call-center representative making a split-second
decision about customer service, the information an enterprise assembles on virtually every aspect of the organization—customers, prospects, products, inventory, finances, assets, or employees—can have a significant effect on the organization’s ability to satisfy customers, reduce costs, improve productivity, or mitigate risks.

Good data hold the key to better decision making, as they give an organization the background to assess business situations and locate trends that provide a competitive edge. This is a “data payoff” that organizations need to see, as the amount of business is estimated to be doubling every 1.2 years (Lohr 2011). But an organization cannot simply cleanse and improve the quality of its data, and then expect these data to be a static resource over time. Data are a fluid, dynamic, and ever-changing resource, so even the most elegant data quality initiative is not a once-and-done activity (Fisher 2009, 128).

For example, companies acquire new customers, products, inventory, and assets every day. Additionally, companies regularly add new data sources from new trading partners in the supply chain—or they add new customers through M&A activity or organic growth. This dynamic data environment can lead to a number of adverse business impacts, including fraudulent disbursements, undue risk exposure, payroll overpayments, and underbilling (Loshin 2011, 7). To avoid these potential problems, a more expansive and repeatable data quality strategy is necessary, especially in information-driven organizations.

CASE STUDY: DATA MANAGEMENT IN PRACTICE

This article will examine the use of data management technologies by iJET International to better investigate and control a large and diverse set of travel information. iJET is an intelligence-driven provider of business resiliency and risk management solutions to nearly 500 multinational corporations and governments, helping the organization prepare for and respond to global threats to its people, facilities, and supply chain assets. iJET’s Worldcue Global Control Center technology solutions give decision makers and organizations the real-time information needed to anticipate and respond to business disruptions with a competitive edge. iJET’s Worldcue solutions provide decision makers with actionable information on potential disruptions and emerging threats to employees, operating assets, and suppliers.

iJET is an interesting use case for data management due to the amount of data it amasses on travelers and trips each year. Along with data on client operating assets and suppliers, iJET tracks more than 14 million client trips each year. Information on these trips is received from thousands of different sources, such as employees, travel systems, and travel agencies. Due to the close relationship between its information and the services it provides, iJET is an information-driven company. While manufacturing companies cultivate and deliver products, or pharmaceutical companies research and refine drugs and therapeutics, iJET’s business model focuses on providing high-quality data on the travel process as well as dedicated services and support built on top of this information framework.

To serve its network of organizations, iJET takes an employee-centric model. Worldcue Traveler allows employees to log in and view their current and recent trips. Employees, corporate travel managers, and corporate security managers use this information for employee tracking, tax records, planning, and so on.

Travel information is notoriously tricky, and iJET faces an interesting mix of issues to harmonize and synchronize this wealth of information. More than 50 intelligence analysts and subject-matter experts have access to 15,000 sources of intelligence that track global threats that may impact employees, expatriates, assets, and suppliers. The data flowing into iJET can come from more than 2500 data sources.

Outside of the sheer size and scope of the data that accumulate during millions of trips, the unique aspects of travel data pose their own set of issues:

- Trips contain, on average, 3.8 segments (or legs). So, each trip is actually composed of several subcomponents that belong to a master data element.
- Due to the number of times trips change and the number of data sources iJET uses, information on the same trip can arrive at iJET between five and 10 times.

The data are complex not just in the number of elements related to a single trip (for example, multiple airline segments, hotel stays, rental cars), iJET also faces a high degree of duplicate or redundant data
elements. These characteristics lead to four data management phases for iJET:

1) **Analyze the integrity of information through data profiling.** Also known as data discovery, data profiling provides an overview of the relative strengths and weaknesses of an organization's data. This sets the stage for efforts to improve the quality of information later in the process.

2) **Manage inconsistent, incomplete, or incorrect data.** Relatively minor issues, like nonstandard phone number patterns (919.555.1212 vs. 9195551212), often lead to data challenges when trying to analyze data trends. In addition, missing fields or partial information complicate efforts to provide comprehensive service, while also presenting a challenge for downstream data analysis efforts.

3) **Resolve duplicate records.** Multiple instances of employees' profiles or trips sourced from multiple systems make it difficult to provide an accurate picture of the risk associated with a threat as well as make accurate decisions to mitigate the risk associated with the threat. Resolving these duplicate records is the primary objective of iJET's data program.

4) **Conduct ongoing data monitoring and analysis.** Building off the previous stages, it's important to view data management as a journey, not a destination. This phase acknowledges that data quality can become part of the corporate culture, often known as "data governance."

**UNDERSTANDING DATA CHALLENGES THROUGH DATA PROFILING**

iJET is in a position familiar to many enterprises; it has little-to-no authority over the source of a client's data. For organizations that buy data from third-party sources or receive data from suppliers or distributors, this is a common position that leads to some specific challenges. Often, there is no established set of standards for how an employee will be represented across two different systems. In fact, the representation of a single employee may vary even within the same data source.

The presence and proliferation of nonstandard or poor-quality data can have an immediate impact on iJET’s ability to serve its client base. If an employee has an incorrect email or phone number, it will obviously be more difficult to reach that individual in an emergency. If employees are assigned to the wrong client or business unit, iJET will have more difficulty responding to blanket requests from the organization for details on the travel status of a group of employees. Without a standard way to reconcile that data, these errors can happen repeatedly, putting undue strain on both the IT and business groups tasked with managing data.

The first step for iJET—and for any organization that is undertaking a data management initiative—is to conduct an accurate assessment of the data it has as well as new data sources as they become part of the corporate information environment. This phase of data discovery provides a blueprint of the logical steps of a data management initiative, and it can become an integral part of any ongoing data management practice.

Data profiling techniques generally fall into three groups:

- **Structure discovery:** Do the patterns of the data match expected patterns? Do the data adhere to appropriate uniqueness and null value rules?
- **Content discovery:** Are the data complete? Are they accurate? Do they contain information that is easily understood and unambiguous?
- **Relationship discovery:** Do the data adhere to specified required key relationships across columns and tables? Are there inferred relationships across columns, tables, or databases? Are there redundant data?

The first type of data profiling, structure discovery, examines the basic characteristics of the data—their inherent arrangement and organization—to gauge if they are a good fit for the intended purpose. A prime example is pattern matching, which determines whether the data values in a field are in the expected format. For example, pattern matching analyzes whether a phone number field contains all phone numbers. Pattern matching also uncovers whether a field is all numeric, if a field has consistent lengths, and other format-specific information about the data.

Consider a pattern report for North American phone numbers. There are many valid phone number formats, but all valid formats consist of three sets of numbers...
(three digits for area code, three digits for exchange, and four digits for station). These sets of numbers may or may not be separated by a space or special character. Pattern analysis would identify the following patterns as indicative of a phone number:

- 9999999999
- (999) 999-9999
- 999-999-9999

Content discovery, the next type of data profiling, goes beyond the structure of the data and evaluates the meaning of the information. Content discovery techniques can uncover nonstandard data or outliers to find data elements that don’t make sense.

For data elements like organizational names, content discovery can uncover similarities and differences within or across data sources. An organization’s name can be represented in a variety of different ways. For example, “GM,” “General Motors,” “G.M.,” “g.m.,” and “Genrl Mtrs” could represent the same company. Other data elements, such as addresses, are also prime candidates for content discovery. In a data source, “100 E 4th Str,” “100 East Fourth Str.,” “100 East Fourth,” and “100 4th Street” can all represent the same address.

The final data profiling technique, relationship discovery, links data in disparate applications based on their relationships to each other—or to a new application being developed. Different pieces of relevant data spread across many individual data stores make it difficult to develop a complete understanding of the data. Relationship discovery helps understand how data sources interact with other data sources.

For example, a customer ID, phone number, or email address may serve as the primary identifier in a data source. However, a sales order—or in iJET’s case, a travel ticket—can exist that has no corresponding customer in a client database. Alternately, a customer database can have multiple customer records with the same ID. These profiling reports uncover areas where duplicate or potentially invalid data occurs.

For iJET, these data profiling reports are an ongoing part of the data management routine. As new data enter the organization, these reports provide an understanding of the relative health of these data—and what iJET must do to make the data more useful. After profiling the data, the company can move to the next few phases, which focus on improving the quality of that information.

### FIGURE 1

The rules for a parsing engine can provide multiple resolutions to a text string and provide a scoring to validate the likelihood of a successful solution.

#### ADDING VALUE TO INCONSISTENT, INCOMPLETE, OR INCORRECT DATA

A necessary foundation for any data management effort is to examine data at their most elemental level, establish patterns according to a repeatable nomenclature, and begin the “dirty work” of cleaning data. From the data profiling phase, iJET gained insight into the key areas to address during a data improvement process. A necessary element of this process is parsing, or separating data into usable components.

For a given data type, a data quality engine uses customized pattern libraries that are tuned for the specific type of content being processed. These definitions include pattern rules, word vocabularies, and other processes that allow the engine to intelligently classify raw strings of data into potential tokens. After categorizing and identifying every token in a field, the engine identifies the best pattern match based on heuristic rules. This allows for meaningful data elements to be derived from raw string data.

The image in Figure 1 represents one of the processes used to parse a name into multiple patterns. This is a parsing algorithm designed to break up the components of a rather simple data element—an individual’s name. In Figure 1, the parsing engine identifies four potential solutions (indicated by the “N of 4” text next to the score) for breaking the name into first name, middle name, and family name. The engine then selects the parse solution with the highest score and outputs the name tokens according to the defined pattern.

The example illustrates the complex logic applied during the parsing process. The first pattern, solution 1,
matches the name to a known pattern [Given Name] [Middle Name] [Family Name]. Solution 2 matches the name to a different pattern [Family Name] [Given Name] [Middle Name]. In this case, the solution is given a lower score by the engine because the pattern is less likely to occur than the previous pattern. Solution 3 is also scored as low-likelihood patterns, so the natural language engine decides that solution 1 is the best way to break apart the data (the “best solution”). The result is an automated way to break up text into its most useful components. Applied to a data set with different name formats, these rules would parse data into the elements depicted in Table 1.

Parsing techniques allow iJET to derive usable data elements from inconsistently formatted raw data. By applying parsing early in the process, iJET simplifies downstream systems that consume the incoming data.

Standardization routines build on the work of parsing and begin to find ways to standardize or normalize each component. In this way, technology can eliminate semantic differences found in source data, including multiple spellings of the same information, multiple patterns of the same data, or the translation of inventory codes to product descriptions.

One type of standardization is “phrase standardization,” which describes the process for modifying original source data from some value to a common value. The company name “GM” can be transformed based on standardization rules from a variety of original values (G.M. Auto or Gen’l Motors) to General Motors. This will be the standard view of this company, and a technology can use predefined rules to make these connections; they can be imported from other corporate information sources (as a glossary of predefined rules or lookup tables).

While phrase standardization looks at elements of the phrase itself, pattern standardization includes the decomposition of source data into atomic elements such as first name, middle name, and family name. After identifying the atomic elements, pattern standardization reassembles each token into a common format. This builds on the approach given in the parsing section, adding a new wrinkle: the ability to resolve seemingly small, but meaningful, differences in the data elements. This can include multiple name formats such as [Name Prefix] [Family Name] [First Name] [Name Suffix] and [Family Name] [First Name].

Using a rules-based approach, a data quality engine can identify the parsed element’s first name, family name, name prefix, and other potential components of the name such as professional designation and name suffix. The standardization engine then reassembles the tokens into a common pattern such as [Name Prefix] [First Name] [Middle Name] [Family Name].

Another standardization technique, element standardization, involves the mapping of specific elements or words within a field to a new, normalized value. The following example illustrates how the engine is able to isolate a specific word such as “1st” and then modify the single word into a standardized word such as “First.”

While this all can seem rather complex, data quality technology allows organizations to approach standardization more holistically. Figure 2 illustrates the automatic discovery of semantic differences found in a company name field. In this case, the engine is able to automatically identify the semantic differences, pattern differences, and other common inconsistencies. The engine groups similar data into a single group and then builds rules to normalize each entry to the most frequently occurring entry. This advanced analysis allows business and IT users to rapidly create custom data correction rules based on a thorough analysis of the source data.

The left frame displays automatically discovered semantic issues grouped together, while the right frame displays automated mappings. If a user wants to change a standardized value (for example, “Ford Motor Company” instead of “Ford Motor”), this console provides that capability. Not only is the standardization scheme changed here, but subsequent runs of this rule will apply the new rule to the new data going through the process.

iJET applied various standardization techniques to its data to make them more presentable to the customer. This allowed iJET to refer to its employees in a consistent manner across each trip they take.
MATCHING AND RESOLVING DUPLICATE DATA

A primary component of the iJET data management strategy was to identify duplicate records and rationalize information into one master record. It is common, for example, to receive trip information from a corporate credit card that does not have a hotel listed—and separate information from a travel agency with the hotel record included. Before the data management effort, up to 10 percent of their records were duplicates that had to be resolved by human interaction. If a client needed to know how many employees were in London on a given day, for example, the number could have been off by 10 percent because an employee record might appear more than once.

The effect of duplicate data can have a material impact on a business like iJET. During a regional or national emergency, such as political unrest in Egypt or an earthquake in Japan, iJET must rapidly pull information on the employees in these regions and begin to make plans for alternate accommodations and travel schedules. If there are duplicate data on these individuals, iJET may not know if it has 10 client employees in these regions—or 8, 9, 11, or 12. This is just one way duplicate data can affect iJET’s ability to service the client, but it is a very high-visibility rationale for data management efforts.

To address this issue, iJET focused on a strategy commonly known as data matching or “entity resolution.” Entity resolution is the process of merging multiple data sources (or duplicate records within a single data source) in such a way that records referring to the same physical object are treated as a single record. Records are matched based on the information they have in common. The merged records appear to be different, but can actually refer to the same person or thing. Entity resolution is typically used to cluster similar records (for example, members of the same household) or eliminate duplicate records (Loshin 2011, 295).

A rules-based matching engine provides a combination of parsing rules, standardization rules, phonetic matching, and token-based weighting to strip the ambiguity out of source information. After applying hundreds of thousands of rules to each and every field, the engine outputs a “match key”—an accurate representation of all versions of the same data generated at any point in time.

Because not all data sources are created equally—and not all data elements have the same type or number of attributes—the matching process must allow some variability. A “sensitivity” setting within a matching technology allows the user to define the closeness of the match in order to support both high and low confidence match sets. For example, a company may have a need to find multiple occurrences of a customer across systems, forwarding those identified matches to an administrator for review. In that case, a “looser” match would pull in more matches for review, providing more human input during the process.

To address this issue, iJET focused on a strategy commonly known as data matching or “entity resolution.” Entity resolution is the process of merging multiple data sources (or duplicate records within a single data source) in such a way that records referring to the same physical object are treated as a single record. Records are matched based on the information they have in common. The merged records appear to be different, but can actually refer to the same person or thing. Entity resolution is typically used to cluster similar records (for example, members of the same household) or eliminate duplicate records (Loshin 2011, 295).

A rules-based matching engine provides a combination of parsing rules, standardization rules, phonetic matching, and token-based weighting to strip the ambiguity out of source information. After applying hundreds of thousands of rules to each and every field, the engine outputs a “match key”—an accurate representation of all versions of the same data generated at any point in time.

Because not all data sources are created equally—and not all data elements have the same type or number of attributes—the matching process must allow some variability. A “sensitivity” setting within a matching technology allows the user to define the closeness of the match in order to support both high and low confidence match sets. For example, a company may have a need to find multiple occurrences of a customer across systems, forwarding those identified matches to an administrator for review. In that case, a “looser” match would pull in more matches for review, providing more human input during the process.

As well as offering predefined match rules, it is important to have a configurable or customizable match engine that allows an organization to define unique or highly tailored matching rules. These matching rules can include any number of fields as well as any number of match conditions coupled together with Boolean rules (AND/OR). Matching records are then assigned a single group ID, which can be persisted and maintained over time. This also allows users to group the duplicate records based on linkage rules or to automatically consolidate duplicates into a single “best record.”
A matching engine can remove ambiguity from an address and generate a unique key for three distinct representations of the same address. A rather simple address—100 North Main Street—can have various permutations across data sources: 100 N Mane St Floor 12; 100-12 North Maine; No. Main St. #100 12th.

Using predefined address parsing rules, the match engine can accurately identify the meaning of each token within the string. A token, in this case, would be each individual element of the address. A match engine can identify the different elements of common data types, such as a physical address in the United States. In this case, the engine breaks the address into the following tokens (see Table 2).

After eliminating pattern differences, the match engine then applies a series of manipulations to each token and, after applying hundreds of thousands of rules, outputs a unique, system-generated match key. The match key can be used to identify the address relationship across any number of disparate data sources and, by combining other fields such as business name and postal code (such as a ZIP code) into the match process, the records can be rationalized as a single entity.

Following this stage, the engine can do one or all of the following:

1) Display the matches in report format
2) Automatically consolidate the records into one “best record”
3) Append grouping keys to each record and persist the keys over time
4) Write match keys to an index or a cross-references table

For iJET, the types of duplicate data could span a variety of fields. There could be duplicate employee information such as name, address, phone number, or email address. The individual travel segments can have a number of duplicate data points from different data sources. Each of these elements was a prime candidate for entity resolution, and allowed the company to reduce duplicate data significantly during the initial stages of the implementation.

### ONGOING DATA MONITORING AND ANALYSIS

For iJET, the clean-up process was part of an ongoing data governance initiative. Data governance is an enterprisewide initiative that acknowledges the inherent value of corporate information and provides a framework of people, processes, and technology to manage these data over time. For an ongoing data governance effort, it is critical to get feedback on the health of corporate information, particularly as it provides a link between the overall integrity of data to the ability of data quality team members to resolve issues within an acceptable timeframe (Loshin 2011, 237).

The iJET data governance program, outlined in Figure 3, shows how data governance pushes beyond the data itself and acknowledges that high-quality, reliable information requires the attention of a host of people across a variety of business processes. The data governance effort required adjustments in the employees who managed data, as well as a change in the way IT systems consumed and managed data. These alterations in the IT landscape support a stated business goal, allowing iJET to verify that the project achieved a measurable business impact.

Data monitoring technology gives organizations the tools they need to understand how and when data strays from its intended purpose. Monitoring also helps identify and correct these inefficiencies through the automated, ongoing enforcement of customizable business rules. Data monitoring ensures that once data become consistent, accurate, and reliable, they remain that way, giving confidence to professionals who make decisions in the organization.

For iJET and other information-based companies, the questions answered by data monitoring are fundamental to the business itself. Are the data still valid? Are they still meeting their intended use? Do they represent the processes accurately? Are there violations in these processes? Which groups are creating the most nonstandard data elements? A data monitoring methodology and technology provides greater control over data, giving business analysts and IT staff constant and repeatable feedback on the quality of information in core systems.

### TABLE 2

Multiple variants of an address parsed into subcomponents

<table>
<thead>
<tr>
<th>Street Number</th>
<th>Pre-Direction</th>
<th>Street Name</th>
<th>Street Type</th>
<th>Address Extension</th>
<th>Address Extension Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>N</td>
<td>Mane</td>
<td>St</td>
<td>Floor</td>
<td>12</td>
</tr>
<tr>
<td>100</td>
<td>North</td>
<td>Main</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>100</td>
<td>No.</td>
<td>Main</td>
<td>#</td>
<td></td>
<td>12</td>
</tr>
</tbody>
</table>

©2011, ASQ
A data monitoring solution can help maintain consistent, accurate, and reliable data by:

- Identifying trends in data quality metrics and data values
- Providing instant alerts for violations of pre-established business rules
- Quantifying the costs associated with data quality and business rule violations
- Detecting variances from cyclical runs
- Recognizing when data exceed preset limits, allowing an administrator or analyst to immediately update these data by addressing the problems upfront before the quality of data declines

A key element of data monitoring, particularly in regard to cross-functional programs like data governance, is its utility across IT (technology savvy) and business (process savvy). The IT group can work in conjunction with business staff to install rules on the data coming into the organization or that are being added to existing applications. During this process, the groups will work together to establish business rules that govern what is “good” and what is “bad” at the data level. The business rules include detecting unexpected variances, validating numeric calculations, testing domain values against a list of acceptable values, or validating data before allowing a transaction to continue.

Once the technology discovers violations, the data monitoring software notifies other designated users of rule violations and/or corrects the data. The monitoring engine can copy the violated data to a repository for further examination and correction. The ability to create a number of different outcomes (for example, write to a file, create an alert, start a corrective routine, and so on) is extremely powerful, as it gives users the flexibility to create complex business rules to manage virtually any situation.

With these capabilities, the business rules engine becomes a tool that ensures data meet organizational standards. In doing so, these rules become a critical control mechanism within data governance or operational compliance efforts.

For any company attempting to gain more control over enterprise data, data monitoring provides the ability to report the changing nature of corporate information over time. The data monitoring can start with an initial data profiling before a data quality improvement program to provide a baseline of data integrity. Ongoing data monitoring throughout different parts of the process gives more detail on the progress of the effort, finalizing with a set of data monitoring controls, often displayed as dashboards. Figure 4 demonstrates a dashboard screen, showing the number of triggers or violations experienced by certain business rules.

This dashboard provides the type of information a manager or an executive needs to understand the state of data within the enterprise. These web-based reports provide a baseline of knowledge about any problematic data.

The business analyst or data steward working directly on data quality initiatives would have similar, but more granular, screens that provide the understanding required to manage enterprise-wide data quality initiatives. These dashboards would fuel much of the work done by these employees, providing an automatic way of managing data quality over time.
For iJET, the outcomes of these data monitoring reports provide a unique method of feedback to the sources of its data. The organization performs a weekly analysis of data trends to see which of its clients has the best and worst scores relative to data quality, including a score before and after the clean-up process. This information is presented to the client organization—and internally at iJET—to help everyone understand the value of the information coming in, as well as ways the team could improve data quality metrics over time.

LESSONS LEARNED FROM THE DATA QUALITY FRONTLINES

iJET is in a unique position from a data management perspective. As a voracious consumer and organizer of vast amounts of data, the company must have the highest level of controls on this data to remain an effective, customer-focused organization. After a multiyear data quality program, the organization learned several lessons that can be applied to all organizations.

First, iJET learned early on that data were more than an IT problem. While many organizations regard data as an output of business systems, which are the domain of the IT group, iJET viewed data as an asset that could be managed similarly to a building or a truck. Just as the maintenance on a truck or a piece of production equipment has defined steps and processes, iJET has a set of procedures for the collection and management of data. To accomplish this, it established teams and formal processes to go along with data management technology, creating a foundation for a progressive data governance program.

A second takeaway was the understanding that a data management initiative is a momentous challenge, and it is a best practice to provide a dedicated plan and set of goals for the program. The team soon learned that measuring and quantifying known issues (for example, finding the origin of bad data or understanding the extent of the problem) was just as important as cleaning up poor-quality
data. This realization allowed iJET to mitigate risks in the short term by understanding the utility of some data sources while providing a long-term vision of where the company could go with its data governance program.

Finally, iJET understood that a culture of data governance helped it evangelize the need for better data to its client. By communicating more effectively to the client base (via data monitoring reports), iJET was able to show how improvements in data could lead to improved services for the employee. This tie between the quality of data and a service level was new in the minds of many clients, but repeated and consistent feedback helped establish this link.

REFERENCES


Thanks to Reviewers
In addition to the members of the Editorial Review Board, whose names are listed in each issue, each submission to SQP is examined by up to three of our distinguished reviewers. These professionals contribute their evaluations and recommendations for the SQP manuscripts anonymously as part of our double-blind review process. Once a year, as we reach the end of a volume, SQP wishes to acknowledge the contribution of those individuals who reviewed submissions in the past year. For volume 13, 2010–2011, those contributors were:

Fernando Brito e Abreu
Jonathan Addleston
John Arce
Steven Arndt
George Besseris
J. David Blaine
Robert Charette
Leo Clark
Tom Ehrhorn
Uma Ferrell
Tom Flynn
Christopher Fox
Eva Freund
Karol Früehauf
Trudy Howles

Leonard Jowers
Derek Kozikowski
Michael Kress
Carla Langda-Sivak
Gloria Leman
Shin Liu
Paulo Lopes
Rupa Mahanti
Varun Malik
Philip Marriott
Joe McConnell
Linda Melnnis
Denis Meredith
Vivek Nanda
Mark Paulk

Jason Pryde
Nicole Radziwill
Steven Rakitin
Prem Ranganath
Henry Schneider
Stephen Sheng
Jim Sivak
Malcolm Stiefel
John Suzuki
Louise Tamres
Larry Thomas
Unnikrishnan Valliyil
David Zubrow

BIOGRAPHIES
Brett Dorr is senior director of product management for DataFlux Corporation. He joined DataFlux in 2003 and directs the company’s product management team. In this role, he guides the company’s research and development team to provide innovative data quality, data integration, and master data management solutions for a broad customer base. Previously, he was the director of sales engineering at DataFlux, working closely with customers to design effective solution architectures in pre- and post-sales consulting engagements. Prior to joining DataFlux, Dorr held product management and development roles at SAS, the parent company of DataFlux, and helped develop products such as SAS ETL Studio (now SAS Data Integration Studio), SAS Management Console, and SAS for Windows. Dorr holds a bachelor’s degree in information systems from Kansas State University. He can be reached at brett.dorr@dataflux.com.

Rich Murnane is a data architect and the director of the iJET Enterprise Data team. Murnane and his team are responsible for the collection, processing, monitoring, and quality of client and iJET production data. He has 15 years of experience designing, building, and supporting database systems. He is an active member of the Data Administrator Management Association and Mid-Atlantic Oracle User Group. He is a founding member of the iJET Professional Services team, which provides travel industry and data management services for corporate and federal government clients. As an active member of the data management community, Murnane has been blogging about data management for more than five years and is also a member of the DataFlux Client Advisory Board. He holds a bachelor of science degree from Salisbury University with concentrations in computer science and mathematics.